Faculdade de Engenharia da Universidade do Porto

Inferring Urban Indicators Through Computer Vision on Google Street View

Sara Isabel Linhas Paiva

Dissertation

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: Rosaldo J. F. Rossetti, Phd

Co-Supervisor: Diogo Santos, MSc

February 28, 2018

Inferring Urban Indicators Through Computer Vision on Google Street View

Sara Isabel Linhas Paiva

Mestrado Integrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

Chair: Prof. Dr. Rui Carlos Camacho de Sousa Ferreira da Silva
External Examiner: Prof. Dr. Brígida Mónica Teixeira de Faria
Supervisor: Prof. Dr. Rosaldo J. F. Rossetti

February 28, 2018

Abstract

The physical environment of a community has been proven to have effects on the mental and physical state of a population. As such, the extraction of Urban Indicators (UI) that evaluate the effects of urban development is essential to assert relationships between the surrounding environment and the well-being of a society. Such a relationship, for example, would be the role of green areas in a city on the prevalence of obesity in its population. Indeed, large green areas may suggest that people spend some quality time outdoors, doing some physical activities. In addition, these indicators can contribute to the identification and preventive action in risk situations. For instance, a very degraded area with too much waste accumulated may pose serious risks to public health.

However, the traditional methods for UI extraction, particularly in the case of physical indicators, are limited due to the lack of standardized data organization, the subjectivity of self-reported responses, while generally being highly resource intensive and costly. This dissertation aims to apply Computer Vision to provide a means to automate the extraction of UI, overcoming the limitations of the traditional approaches, by taking advantage of tools that offer remote visualization of locations at low cost. The success of this approach though, depends on the accurate identification of physical Urban Indicators that can be extracted from an image, and on choosing appropriate Computer Vision techniques to provide the most precise results for this analysis. The proposed solution is a platform able to process and classify images according to Urban Indicators in a given location. A proof of concept for such an approach involves the selection of points of interest for analysis in an area according to a set of criteria defined *a priori*. Images are then extracted at the selected coordinates, after which results are presented.

The expected contribution of this dissertation is to offer an effective approach to infer visual urban indicators in a consistent format, demonstrating that Computer Vision applied to feature extraction and image classification has sufficiently evolved to the point of being able to automate this type of analysis, thus representing a sustainable alternative to traditional methods. A literature review is presented, which covers the main concepts related to feature extraction, as well as are discussed some related work. The final product of the dissertation was capable of extracting urban indicators from a given area; however, it lacks robustness in terms of the number of different features the approach is able to extract. Therefore, more suitable datasets for training the algorithm are necessary to further improve the proposed solution herein discussed.

Keywords: *Computer Vision, Visual Semantics, Urban Indicators, Omnidirectional Streetscape Imagery, Google Street View.*

Resumo

Está provado que o ambiente de uma comunidade afeta o estado físico e mental de uma população. Portanto, a extração de Indicadores Urbanos (IU) que avaliem os efeitos do desenvolvimento urbano é essencial para estabelecer relações entre o ambiente circundante e o bem-estar de uma sociedade. Uma relação possível seria, por exemplo, o papel de áreas verdes numa cidade e a prevalência de obesidade na sua população. De facto, espaços verdes extensos poderão sugerir que a população gaste tempo ao ar livre, fazendo atividades físicas. Adicionalmente, estes indicadores podem contribuir para a identificação e tomada de ações preventivas em situações de risco. A título de exemplo, uma área muito degradada com resíduos acumulados pode representar riscos sérios à saúde pública.

No entanto, os métodos tradicionais de extração de IU, particularmente no caso de indicadores físicos, estão limitados devido à falta de organização de informação padronizada, à subjetividade de autorrelatos, tendo de forma geral custos elevados a nível de recursos. Esta dissertação tem como objetivo aplicar técnicas de Visão por Computador para providenciar resultados mais precisos a esta análise. A solução proposta é uma plataforma capaz de processar e classificar imagens de acordo com Indicadores Urbanos numa certa região. Uma prova de conceito dessa abordagem envolveu a escolha de pontos de interesse para analisar numa área, de acordo com um conjunto de características definidas *a priori*. As imagens são depois extraídas nas coordenadas selecionadas, sendo os resultados apresentados posteriormente.

A contribuição esperada desta dissertação é oferecer uma abordagem eficiente para inferir indicadores visuais físicos num formato consistente, demonstrando que a Visão por Computador aplicada à extração de características e classificação de imagens evoluiu ao ponto de ser possível automatizar este tipo de análise, representando assim uma alternativa sustentável a métodos tradicionais. É apresentada a revisão bibliográfica cobrindo os conceitos principais relacionados com a extração de características, bem como discutindo trabalho relacionado. O produto final da dissertação é capaz de extrair os indicadores urbanos de uma região escolhida; no entanto, falha na robustez em termos do número de características diferentes que a abordagem consegue extrair. São, portanto, necessários conjuntos de dados mais adequados para treinar o algoritmo e assim melhorar a proposta aqui discutida.
Palavras-chave: *Visão por Computador, Semântica Visual, Indicadores Urbanos, Imagens de Rua Omnidirecionais, Google Street View.*

Acknowledgements

I want to thank my supervisor, Professor Rosaldo Rossetti, for the accurate and useful input. More, for wholeheartedly supporting me through changes of the project and never saying that it could not be done. To my co-supervisor and good friend, Diogo Santos, who first proposed this theme: thank you for your constant availability to discuss the limits and extensions of this project. You kept me in the right track and in the track. Plus, it was no small feat to survive being the first to go through each extensive sentence of this document. I will get better.

To my parents, for giving me a safe haven when I needed the most, there is no acknowledgment big enough for that. Also to my older brother, sister, siblings-in-law and progeny, for constantly challenging me in humorous ways and for never failing to be present.

To my colleges from the University, and to everyone, from several areas, to whom I asked for input while advancing through the project. A special thanks to Ivo, Teresa, Guillaume, Cris, Artur, Catarina and Alexandre, whose direct talks resulted in some new ideas for this and future projects.

And to all my friend that without doubt lighted a lot of candles in the hopes I would finish the dissertation, thank you. It looks like it worked.

Sara Linhas Paiva

“He saith unto them, Come and see.”

John 1:39

Contents

List of Figures

List of Tables

Abbreviations

AI	Artificial intelligence
API	Application programming interface
CCTV	Closed circuit television
GSV	Google street view
FAST	Features from accelerated segment test
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
PSR	Pressure State Response
RQ	Research question
SIFT	Scale-invariant feature transform
SSD	Sustainable structural design
SURF	Speeded-up robust features
SUSAN	Smallest univalue segment assimilating nucleus
SVM	Support vector machine

Chapter 1

Introduction

1.1 Scope

As urban areas continue to expand and provoking serious damages to the environment all around the globe [YSSW17] there is a need to assure that this growth is done effectively and in a sustainable way. This means that there is a need to identify and correct situations where that growth was not made in a positive manner. In the particular scope of this dissertation, it is noticed that the visual physical environment of a community has effects on the mental and physical state, as well as on the attitude of a population [QD14]. As such, there are cases where a visual identification of issues might help mitigating urban problems. Such is the case of Abuja in Nigeria, where the construction of low-income housing resulted in a high use of energy per household, due to efforts to reduce the high temperatures felt in the buildings. Visual indicators, such as the facade orientation or the existence of urban components that could provide shading, helped identifying and finding solutions for this problem, and its correction in future houses is estimated to be able to reduce between 4 to 29% of energy use [AAMC18].

More than an ideal city far in the future, it is essential to promote the development of smart cities capable of combining smart practices (e.g. sensors, digital applications and software) and methods [Fak16] to extract information from urban environments and act upon it.

Therefore, extracting physical urban indicators of a community allows us to analyze relationships between the well-being of a society and the surrounding environment. It also weighs on the capability of public or private institutions to identify risk areas and take preventive actions accordingly. The aim of this dissertation concerns one of such problems, namely the inference of visual urban indicators, and to provide an approach to surpass the current limitations of such automatic extraction via computer vision techniques.

1.2 Problem Statement and Motivation

Considering the importance of urban indicators, and the relationship between structural characteristics associated with physical and social disorder [SR17], there should be accurate means to extract them. However, the traditional methods, especially in

the extraction of visual indicators, suffer from several limitations, some of which as listed below [RBR+11]:

1. **Secondary analysis of archival data sources** – although the access to such sources is becoming more common and it can be very cost-efficient, the specific purposes of the extracted data result in no standard data organization and there is a possibility that the secondary analysis is an inaccurate synthesis of the primary source [MH05];

2. **Survey-based measures** - methods based on surveys or punctual reports of habitants are influenced by the subjectivity of the people involved, for example by suffering from the same source bias. It is possible to ascertain on the validity of each result by performing a parallel control survey, but this increases the cost of this approach substantially [Uni07];

3. **Usage of empirical audit instruments** - wields accurate results, but considering it needs both suitable technologies and trained human employees, it becomes highly resource-intensive and costly. This means it is usually limited to small geographical areas, not being suitable for a big and diverse analysis.

This brings about the need for alternatives to lessen these limitations. A proposed approach is the use of the available technologies that provide remote visualization of location and study the possibility of automating the inference on physical urban indicators [RBR+11]. Using such technologies for virtual audits, with non-automated methods has been observed to be cost effective, providing more security to the researchers while maintaining the quality of the results. In addition, virtual audits were less time consuming when compared to physical ones (115.3 min to 148.5 min [BOW+10]).

Finally, although the computer vision field is still more vulnerable to image distortions or degradations than human vision [GJS+17], it has had major breakthroughs in the last few years, both in the development of new architectures and its implementation in various technologies as well as applications. As such, it is definitely paramount to analyze its potential to leverage the applications previously described.

1.3 Aim and Goals

The aim of this project is to apply computer vision to get automated results when assessing urban indicators. To reach that aim, it is necessary to go through a number of important tasks, such as to test the results, to parametrize data into intuitive structures, and to check how computer vision compares to more traditional methodologies. These constitute the different goals for this dissertation.

1.4 Document Outline

Considering the large scope of this subject, in order to make the work reproducible, the literature review was based on a systematic literature review and so, this report

starts by explaining the methods used to perform a thorough survey of the related literature, as presented in Chapter 2. Then the review continues by describing concepts related to the main research questions, followed by a brief exposition of the computer vision history, as well as of the challenges and methods related to the extraction of visual semantics from an image. Then the review finishes by presenting a summary of case studies and possible approaches to inferring urban indicators. Chapter 3 presents the main modules for extraction, classification and visualization which underlie the design and development of the proposed platform. The implementation of the modules, the technologies used, the visual representation and analysis of results are discussed in depth in Chapter 4. Chapter 5 remarks on the main considerations and conclusions of this work, emphasizing on what needs to be further improved, and suggests perspectives for future projects originating from this dissertation.

Chapter 2

Literature Review

Since the subject of urban indicators is broad with several definitions, there is a need for a mapping of the current and past knowledge in the field, for which a systematic literature review is essential [Kit04]. So, before inquiring on the proposed approach in terms of image processing and classification, it is important to define what exactly is going to be researched. For this reason, the chapter starts by explaining how the review process was done by identifying the main questions of the issue. A systematic literature review involves a clear defined search strategy, whose goal is to find the most relevant literature possible. It is usually used in broad subjects such as medicine where it is essential to provide an unbiased and thorough analysis. But, with the growth of Software Engineering, there has been a similar need, resulting in the creation of several protocols capable of guiding researchers [MCN$^+$]. Since the results were too extensive, the systematic literature review protocol was only used as a main guideline for this literature review.

Inferring visual indicators using computer vision from an urban environment requires some prior knowledge for both choosing the appropriate indicators of a certain environment, and to understand what is the best approach to have when dealing with remote images acquired using Google Street View [1], or similar services. So it follows that one should start by defining urban indicators and their types of classification so that it is possible to understand what constitutes an indicator in this context. The review continues then by presenting some concepts on computer vision, concluding with the third main topic of the problem, which is how visual semantics extracted by computer vision can be used to assess urban indicators that are related to visual signs.

2.1 Review Approach

2.1.1 Process

The process which the literature review was based upon, systematic literature review, can be viewed in Figure 2.1. The first stage is a general review to assess the main areas of the presented problem.

[1] https://developers.google.com/maps/documentation/javascript/streetview

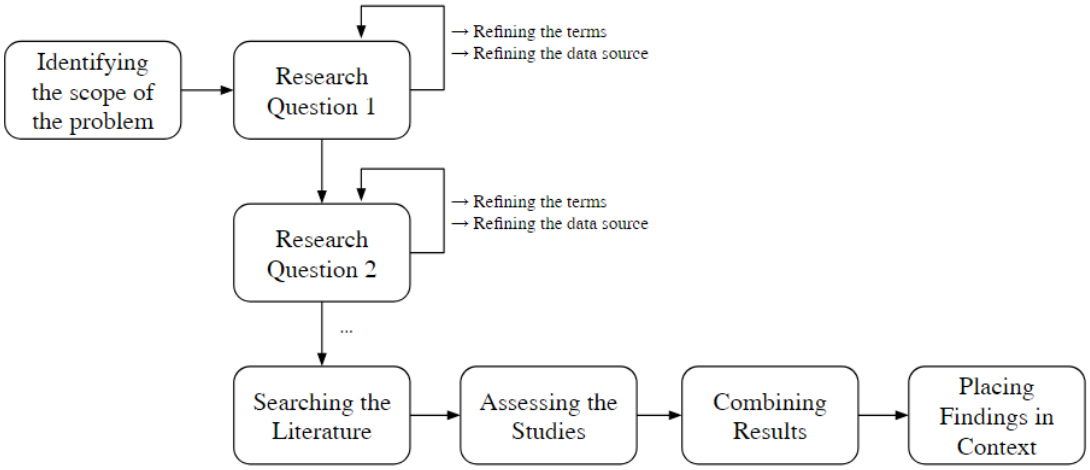


Figure 2.1: Systematic Literature Process

In this section it is explained how the methodology proposed in Figure 2.1 was applied, starting by the research questions definition until the methods. It can be seen in it how in this type of review process the definition of the main issues should be subjected to a detailed peer examination on how the problem is structured. As such, the organization should be improved and restructured when the problem is not clearly presented. To be noted that, because of issues related to bad choice of keywords, the data extraction and qualification wasn't satisfactory, particularly in obtaining precise definitions of concepts.

2.1.2 Questions

The primary research questions of the problem were initially divided in two scopes: the urban indicators and the extraction of information using computer vision. It was later decided that a third area connecting both approaches made sense. Thus, the final division was:

- What are the Visual Urban Indicators that can be extracted?
- How can Computer Vision be used to extract visual semantics from Urban scenes?
- How can visual semantic be used to assess Urban Indicators?

Considering the need to clarify some concepts of the first two questions, subquestions were also defined as seen in Table 2.1.

Table 2.1: Primary and Secondary Research Questions (RQ) List

RQ id	Research Question
RQ1	**What are the Urban Indicators that can be extracted using Computer Vision?**
RQ1.1	What are Urban Indicators?
RQ1.2	What are the Urban Indicators Classifications?
RQ1.3	What are the main issues of Urban Indicators retrieval?
RQ1.4	Retrieval of visual Urban Indicators?
RQ2	**How can Computer Vision be used to extract visual semantics from Urban scenes?**
RQ2.1	What is Computer Vision?
RQ2.2	What is Feature Extraction using Computer Vision?
RQ2.3	What is Image Description using Computer Vision?
RQ2.4	Which Algorithms can be used for Image Description?
RQ2.5	What are the Existing Applications for or using Image Description?
RQ2.6	What are the problems in Feature Extraction from outdoor Environment?
RQ3	**How can visual semantic be used to assess Urban Indicators?**

2.1.3 Methods

2.1.3.1 Data Sources and Search Strategy

Urban indicators are a theme out of the scope of computer science, as such, the data sources had to be chosen in a way that would reflect this. The choice was SCOPUS [2] and ISI Web of Knowledge [3] as primary sources of literature. When more focus on engineering was required, especially when searching for literature related to research question 2, more priority was given to IEEE Xplore [4], ACM Digital Library [5] was also used in this context because of its helpful filtering capabilities. For research question 1, the last data source used was the ASCE (American Society of Civil Engineers) [6] library, indispensable for concepts on Urbanism and urban indicators.

For each research question a general search query, using keywords related to the subject, was defined in order to facilitate the reproducibility of results, as it can be seen in appendix A.

2.1.3.2 Study Selection

Considering the large number of results and the time constraints of this review, not all studies were possible to evaluate. The approach to the study selection consisted in removing duplicates and studies that were published in different sources. Studies published before the year 2000 were also removed, since these were considered outdated, thus of no benefit to the mapping of methodologies and tools related to the extraction of features from images using computer vision. Lastly, studies with little or

[2] https://www.scopus.com/
[3] https://apps.webofknowledge.com
[4] http://ieeexplore.ieee.org
[5] https://dl.acm.org/
[6] https://ascelibrary.org/

Table 2.2: Primary and Secondary Inclusion Criteria (PIC & SIC) and Quality Criteria

Criteria Id	**Criteria**
PIC 1	The study's main concern is the problem presented in the research question
PIC 2	It is a primary study that presents empirical results
SIC 1	The study focuses on specific methods, approaches and/or constraints
SIC 1	The study describes a system, application or algorithm
QC 1	There is a clear statement of the aim of the research
QC 2	The study is put into context of other studies and research

no citations were also discarded. Lastly, studies with less than 10 citations where ignored, unless they seemed particularly relevant after a cursory examination. Another note is that reviews were given priority when searching for concepts (e.g. RQ1.1 and RQ2.1 in Table 2.1).

2.1.3.3 Study Quality Assessment

When assessing the quality of the studies it was necessary to decide on inclusion (primary and secondary) and quality criteria. These criteria are related to the theme of the study and help decide if the work is relevant to the research question being surveyed. It can be seen in Table 2.2 a proposed list of such criteria [KP14].

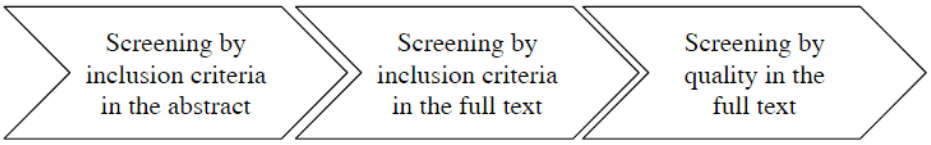


Figure 2.2: Three Stage Filtering of Studies

The criteria are applied to each study in a three stage process as it can be seen in Figure 2.2, from which the end results should be much smaller then the initial set. This process should be documented to clearly state the decisions made during the literature review.

2.1.3.4 General Remarks

Possibly due to a bad choice of keywords, the number of studies returned by each primary and secondary research question (e.g. Table 2.3), made it impossible to analyze the included and excluded studies, in an acceptable time frame.

Considering the structure of the report, the list of non-quantitative main study cases can be seen in some detail in Table 2.4, framed with the rest of the state of the art. Although there was a large variety in terms of approaches to the problem, this made it more difficult to look for specific definitions and concepts.

Table 2.3: Initial number of papers per source per RQ and sub-research question

RQ id	SCOPUS	ACM	IEEE Xplore	ISI Web of Knowledge*	ASCE
RQ1	3	0	341	32	1007
RQ1	3	942	12	32	699
RQ1.1	193	3524	3	4277	4336
RQ1.2	2646	2180	0	3282	9223
RQ1.3	6144	7857	1	6084	2881
RQ1.3	231	20719	1	12914	4151
RQ1.3	1	950	0	16	384
RQ1.4	74	6572	13630	1477	2241
RQ1.4	23	74,658	19	96448	11359
RQ2	4	3993	2	4	-
RQ2	0	568	0	2	-
RQ2.1	1406	287	803	606	-
RQ2.2	9639	1434	15125	1470	-
RQ2.3	173	47	94	58	-
RQ2.4	3	0	3	1	-
RQ2.5	19	731	149	8854	-
RQ2.6	12	77	23	6	-
RQ2.6	18	7	31	9	-
RQ3	0	0	92	0	-

2.2 Urban Indicators

2.2.1 Definition

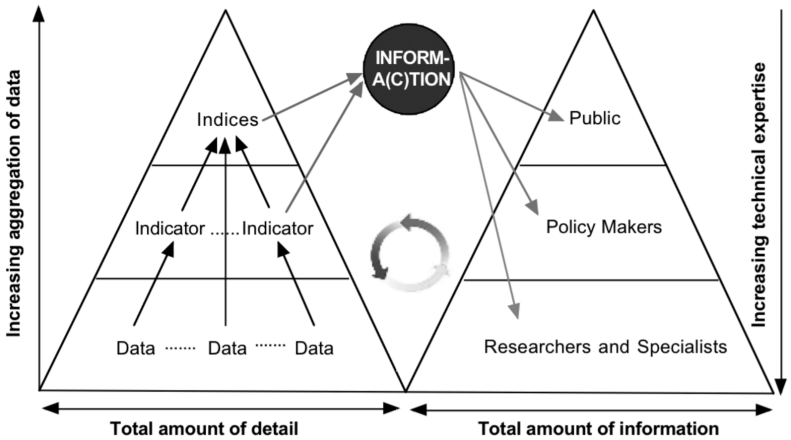


Figure 2.3: Different stages of data aggregation and the application of information according to the technical expertise, from Wu and Wu, 2012 [WW12] based on Braat, 1991 [Bra91]

As Figure 2.3 depicts, urban indicators are an aggregation of data that can be later used to create indexes that facilitate the sharing of information with the public and interested parties. The indexes are, as a result, the formalization of the indicator

aggregating and synthesizing of several data and/or variables. More precisely, using the definition proposed by Gallopin in 1997 [Gal97], a chosen indicator should be capable of inferring the conditions and trends, by itself and when in relation to goals, it should provide an accurate comparison in relation to places and situations providing an early warning and be able to anticipate future trends.

2.2.2 Classification

Taking the previous definition into account when looking for urban indicators (metrics capable of providing information on an urban context), it results in a great variety of possible indicators whose relevance is hard to infer upon. The proposed approach to this problem is choosing urban indicators that are able to provide information on a specific aspect of the environment, such as sustainability.

Table 2.4: Strong Sustainability classification indicators and issues

Indicator	**Definition**	**Issues**
Material and energy flow indicator sets	Analises the urban methabolism by studying the inflows and outflows, for example, the resource consumption and waste production, as well as facilitating the connection between the environment and economy. Base for Life cycle assessment (LCA).	Assessing the life cycle of one component incurs the risk of missing the influence of other components [EG14]
Ecological Footprint (EF)	Measures the amount of land and water needed to provide all materials and energy resources used and to absorb waste in order to support an industry or a population in a certain area.	Demarcation of spatial areas; Does not explicitly consider technology changes or economic aspects [MC12]
Environmental Performance Index (EPI)	Based on two broad policy objectives: protection of human health from environmental harm and protection of ecosystems. Includes 9 areas: agriculture, health impacts, biodiversity and habitat, air quality, climate and energy, sanitation and water, water, fisheries, and forests resources. Each area contributes to the calculation of the Index.	Enconters Insufficient data in indicators. e.g. Freshwater Quality, waste management. [HEL$^+$14]
Green City Index (GCI)	Assesses and compares world cities in terms of their environemntal performance. Mostly based on quantitive data from public sources, varies according to geographical location. An example of Indicators can be seen in Figure 2.4	Enconters issues on the lack of information on certain cites. It has only been taken mostly by the EIU-Siemens project, lack of quality analysis by other researchers. [Eco12]

2.2.2.1 Sustainability Indicators applied to Urban Indicators

Although urban sustainability can be defined with various criteria and different emphasis given to different areas, the general focus is given to the improvement of long-term human wellbeing. This is done by finding balance between the 3 main dimensions generally identified as Economy, Environment and Society, being that different

Table 2.5: European GCI methodology with 16 quantitative and 14 qualitative indicators [Eco12]

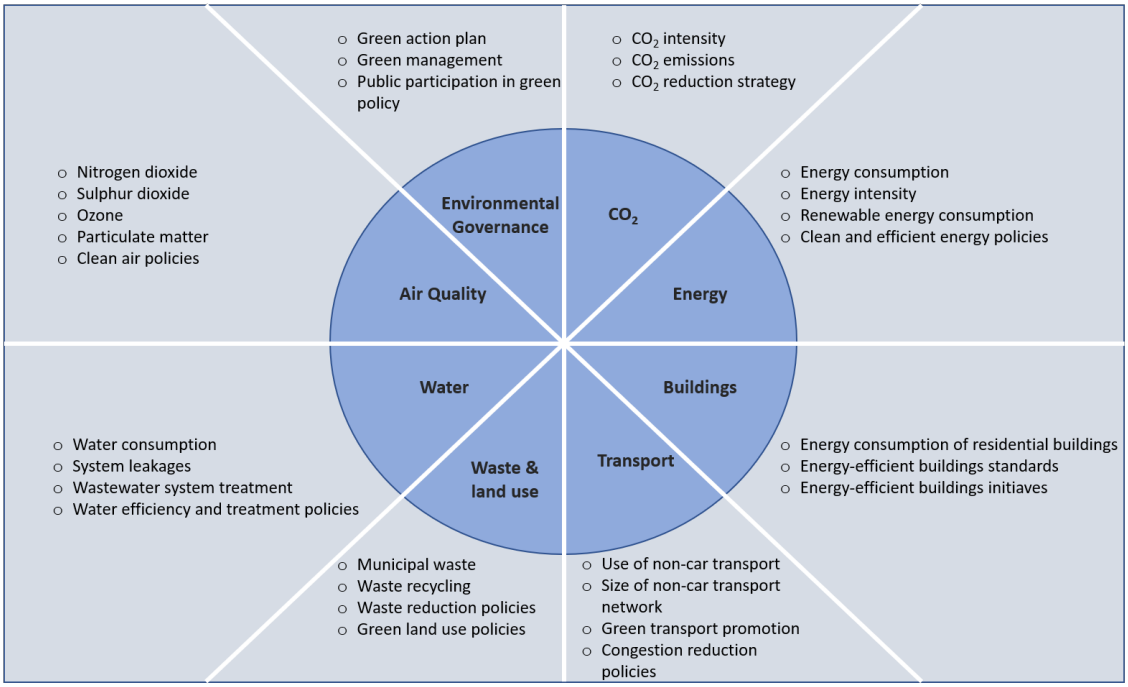


Figure 2.4: European GCI methodology with 16 quantitative and 14 qualitative indicators [Eco12]

sub-dimensions are created with the interaction between these. This classification can be used to frame urban indicators from a sustainability point of view. It has been proposed that an optimal approach to infer what can be a good indicator should classify it according to the Pressure-State-Response (PSR) approach or a framework based in an area/theme. In the PSR approach, indicators are seen as representative of human activities that cause pressure to the surroundings and result in changes to natural resources. These qualitative or quantitative influences usually imply reactions by the respective managing human parties. Besides this it should also be associated to at least one indicator of strong sustainability [HWY15] (that assumes that the natural capital can not be replaced by the human capital. E.g. Renewable energies do not replace the fossil fuels used previously).

For a better understanding of aspects of strong sustainability indicators it is possible to see definitions and issues related to other [HWY15] selected indicators of strong sustainability in an urban environment in Table 2.4.

In Table 2.6 it is possible to observe an example of weak sustainability indicator, with 14 related themes, one example sub-theme [Uni07] and a proposed possible visual indicator to be associated with. In Table 2.7 it is possible to view an example of a strong sustainability indicator applied to a chosen component, air quality [Ole06], with a proposed set of visual components possible to be extracted. The viability of extracting visual semantics from these indicators will be posteriorly analyzed. Both tables allow for the comparison between the different metrics. Weak Sustainability 2.6 analyzes several themes and the existing set of characteristics present without establishing associations between them. In contrast, the Strong Sustainability elements

Table 2.6: Weak Sustainability classification criteria and possible visual urban indicators

Name	**Definition**	Theme	**Sub-theme and Possible Visual Indicators Examples**
Theme-based indicator sets	Organizes according to 4 dimensions of Sustainability: environment, society, economy and instutitions. Sub-organized according to key themes or matters of policy relevance.	Poverty	Living conditions e.g. Prevalence of urban areas classified as slums
		Governance	Crime e.g. Prevalence of areas classified as less safe
		Health	Health status and risks e.g.Prevalence of areas with still water near housing
		Education	Education Level e.g. Prevalence of buildings identified as schools
		Natural Hazards	Vulnerability to natural hazards e.g. Housing near areas classified as hazard prone
		Atmosphere	Air quality e.g. Private vehicles in urban areas
		Demographics	Population e.g. Prevalence of parks for children
		Land	Desertification e.g. Land affected by desertification
		Oceans, seas and coasts	Coastal zone e.g. Housing near coastal areas
		Freshwater	Water quality e.g. Observation on wastewater flow rate in sewer systems
		Biodiversity	Ecosystem e.g. Area of selected key ecosystems
		Economic development	Employment e.g. Areas with prevalence of Income-generating structures
		Global economic partnership	N/A
		Consumption and production patterns	Energy Use e.g. Prevalence of structures for renewable energy .

presented in Table 2.7, display the possible relations and results of the whole process that influences one class of indicator, e.g. Air Quality.

To be noted that, at the time of writing, it has not been found a global classification that would differentiate the visual urban indicators from others. Given this, the proposed approach to this problem is using the existing structures to organize and establish a taxonomy for urban indicators. The main areas in this taxonomy should be based upon the indicators previously studied, such as transportation, green spaces and the external characteristics of the buildings.

Table 2.7: Strong Sustainability classification criteria and possible visual urban indicators

Name	**Definition**	**Indicator Example**	**Indicator Example**	Possible Visual Indicator
PSR-based indicator sets	Organized according to pressures or/and driving forces (primarily anthropogenic processes), system impacts or/and states (current conditions of and impacts on the environment), and responses (societal actions to changes in pressures and systems states).	*Air Quality*		
		Driver	Number of vehicle miles driven Fossil fuel consumption	Number of Private Vehicles identified
		Pressure	Emission of gases e.g. sulfur, monoxide carbon, particulates...	N/A
		State	Ambient air quality for previous pollutants	N/A
		Impact	Materials damage	Signals of degradation in stone structures (e.g.houses and sculptures)
		Response	Incentives to drive less	Identification of parking meters

2.3 Computer Vision

In order to establish an approach capable of a good performance it is important to clarify some concepts on computer vision and visual semantics, proceeding by an exposition of the general methodology in computer vision and most popular algorithms.

2.3.1 Short History

First, to clarify some concepts, **computer vision** is the general term used to refer to the process of extracting information from an image, the science of vision, its definition is usually connected to machine vision, or the study of techniques, methods and the hardware that can be used to construct artificial vision systems for practical applications [Dav12].

Conducting acceptable pattern recognition in pictures using computer processing has been an objective for long. One example is attempting to extract coding information from images like seen in Figure 2.5 using edge tracing and primitive techniques of neuron-like net modules [KA59]. But what is considered to be the most famous pioneering attempt to succeed in computer detection of visual features happened in 1966 and had as an objective the creation of a system capable of "pattern recognition" that would divide an image into regions with likely objects, background areas and chaos [Pap66]. The complexity of the problem was underestimated and it was not possible to complete all the research objectives.

The 1970's marked the appearance of many methodologies, for example using Hough transformation to detect lines and curves [DH72] that would prove to be the foundation of computer vision. During the next decade a more rigorous detection of 3-D objects was developed, resorting to techniques like using the edge contours as a tool for alignment in recognizing occluded objects in cluttered scenes [HU90]. Appearance-based algorithms [VD96, PHD] and later feature-based methods [TR96] marked the 1990s and showed a better understanding of computer calibration.

During the next decade, the importance of getting the general knowledge of the scene [OT01] favored the traditional approach of developing geometric hash functions respective of each model object that would be able to identify that one object in the scene [BGdVL]. The use of the concept of Bag of Words (BoW) [HEHE07, HEH$^+$08] as a means of connecting low-level features to more human understandable high-features for image annotation allowed for the creation of several feature extractors such as SIFT (Scale-Invariant Feature Transform) and SURF (Speeded Up Robust Features) [Tsa12]. The combination of BoW and support vector machines (SVM) was considered the most popular techniques for image classification at the time.

A defining moment of modern-day computer vision is considered to have been the victory in the ILSVRC-2012 competition [RDS$^+$15] of AlexNet [KSH12] which used a deep convolutional neural network to classify the image dataset. It was now possible to take advantage of a method initially presented decades before, since it achieved an error of 15.3%, a great improvement considering that the more traditional runner-up methods showed errors around 10.8% higher. In 2013, a solution similar to AlexNet, albeit using a 7x7 kernel, smaller than the 11x11 approach and thus more accurate, the ZFNet [ZF14] won the challenge again thus cementing the capabilities of Convolutional Neural Networks. In the next year, GooLeNet was able to surpass the human

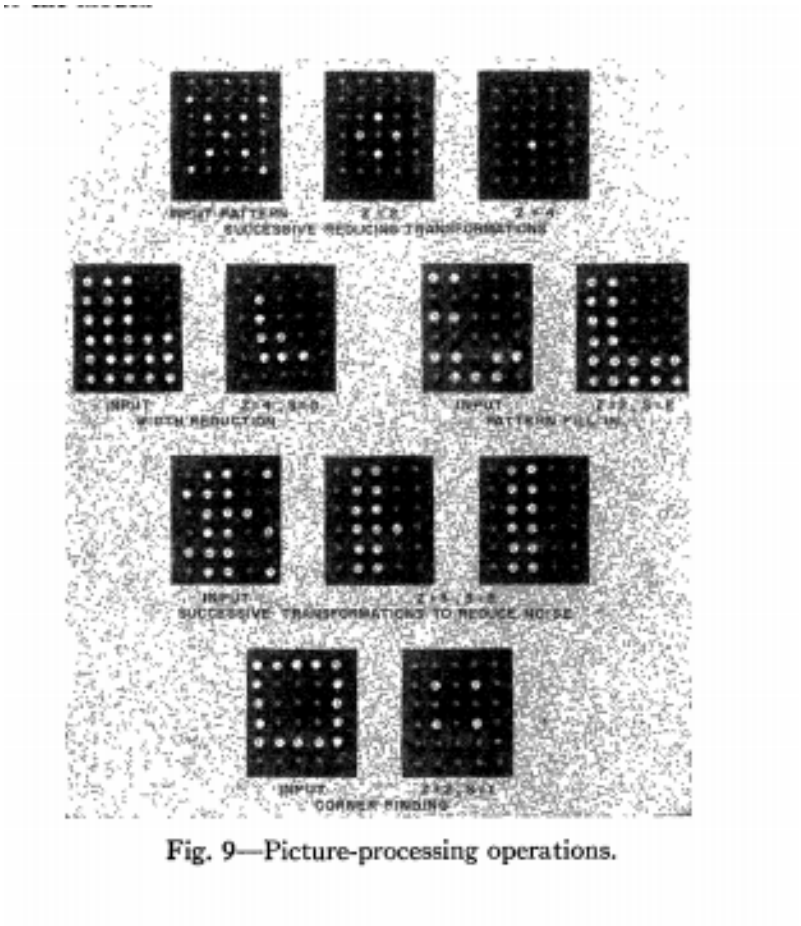


Fig. 9—Picture-processing operations.

Figure 2.5: Basic Picture Processing Operations by Nets of Neuron-Like Elements

classified image set rate which forced the organizers to make another set to be evaluated, this improvement was mainly because of the addiction of a new model with several small convolutions to reduce the number of parameters.

2.3.2 Challenges

To better understand the core of the computer vision methods, one needs to understand the main problems that surround the subject from simple one class object detection to an instantaneous multi-detection in a video.

The main issues of computer vision can be divided in four areas

- Recognition;
- Data Output;
- Segmentation;
- Computation Resources.

The detection and **recognition** of an object is dependent on many issues like, for example, what the representation that should be looked for is. Similar to what was

previously shown, the initial approaches tried to evaluate the object by its 3-D shape and texture, the problem presented was that some objects could be deformable or articulated [KT02]. Adding to this, characteristics vary according to the viewer, the degree of cluttering and general environment of where an object is located. E.g. A traffic sign always has the same shape and color, yet according to the time of the day or the existence of some occlusion it can be hard for a computer to recognize it. Another issue is that while humans perceive that an object might be being used for a certain action, the concept of process of doing an activity to achieve an aim is harder to be understandable by a machine, and that changes the interpretation of a context.

The **data output** is related to how the information extracted from the image is perceived, specially the bridge between low level image features and high level representations. Although low-level features are easy to be extracted from images, it is hard to index that information into classes. The challenge is how to extract high level features, that provide a more conclusive index, in a way that is system effective.

The **segmentation** of an image can be separated into several problems; the most prominent one is that although the models used for the classification of an image where there is only one object present is already efficient, when the image has more classes present and thus more complexity the models tend to fail [AT13]. This raises the question of how an image can be segmented into different parts in a logical manner, meaning the method that can be used to synthesize the amount of information of what is important to extract, taking into account the scale of the different objects for example. Techniques that are used to solve this issue are mainly related to edge and local feature detection/grouping region analysis (identification of color and texture) and motion analysis. Another issue is how the visual learning of an image should be done, i.e. should it look at the image as a whole and then divide it, or should it start by evaluating sections to find small objects so that no detail of the image is lost.

And of course, it should also be **computationally efficient**. This point is especially important considering that the training and testing of models requires datasets of a considerable size and, without them, making the machine learn would not be possible. Vision and comprehension of the world itself is a natural advanced process which humans have not fully grasped, uncertainity still existing about how the image is captured by the eye and processed by the brain through the several years of interacting with the world and classifying objects. As such it makes sense, for now, that an intelligent vision system that is able of going beyond plain conceptual knowledge, thus capable of learning and dealing with realistic contexts, will require a high degree of processing power.

2.3.3 Fundamentals

The interaction between the steps to extract information that allow for an image to be classified can be seen in Figure 2.6. To be noted that the order between segmentation and detection depends on the algorithm chosen. The next part of the report intents to define and give a general approach to the methods used in the steps of computer vision.

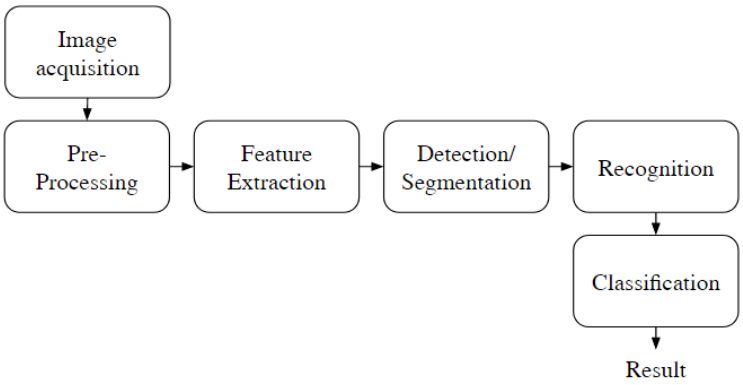


Figure 2.6: Flow diagram summarizing steps of Computer Vision from Image Acquisition to Decision Making based on Davies, 2012 [Dav12]

2.3.3.1 Image Acquisition

Before anything, the quality of the image acquisition is a crucial aspect to diminish as much as possible flaws in the original images that will cause issues in its interpretation and classification. The main aspect to be taken into account is illumination, from the lightning effects to the appearance of highlights and shadows, all can affect how the object is measured and recognized. Techniques to obtain the best results range from applying various sources of light to eliminate the shadows, to more recent hyperspectral imaging approaches (collection and processing across the electromagnetic spectrum [LW04]). Considering that the objective of the present work is inferring urban indicators from images taken from Google street view, these techniques were not considered in depth.

2.3.3.2 Pre-Processing

The pre-processing of an image is the set of low-level processing techniques used to correct minor problems from the image acquisition and to prepare the images for more high level algorithms. Aspects to be taken into account in this phase are:

- Desired type of image (if binary, grayscale or color)
- Basic pixel operations (thresholding, clearing, inverting, copying)
- Brightening of grayscale and contrast-stretching operations
- Noise removal and binary edge location
- Problems around the edge of the image

Most of the operations start with image enhancement through brightness mapping, contrast enhancement, modifications in the histogram, and noise reduction. Mathematical techniques can also be used like convolution, Gaussian filtering, binary dilation and erosion (Figure 2.8b and 2.8c).

Figure 2.7: Original GSV Image [Goo17b]

(a) Gaussian Blur. (b) Erode. (c) Dilatation.

Figure 2.8: Pre-processing image using Gaussian blur, erode and dilate for noise reduction.

(a) Basic Threshold (b) Using Sobel Operator. (c) Using Laplacian Operator.

Figure 2.9: Pre-processing using Threshold, Image Sobel and Laplacian Operator for Edge Detection Example.

An important technique for demarcation of objects in an image is **thresholding**, that broadly consists in separating dark and light regions. One of the main issues associated with this method is finding the optimal value, for which the most often used methods are arbitrary selection or by using the image histogram. An example of thresholding can be seen in Figure 2.9a.

An interesting concept is adaptive thresholding which can be applied to situations when illumination is not uniform in the image by dynamically adapting the values. This can be done by modeling the background, by examining the intensities of pixels in the neighborhood of each pixel to assess on a local threshold. Another way of getting this result is by splitting the image into smaller parts and dealing with each in an independent way.

Filtering is another low-level processing method, especially for noise reduction by application of local mean, median or mode filter. Mainly applied to grayscale images. Options for filtering can be by linear smoothing transformation like Gaussian Smoothing (Figure 2.8a), local averaging or non-linear transformations like median filters and rotating mask.

Although it is necessary some level of smoothing to reduce noise, it is also important to preserve the edge of the objects. Kuwahara, Nagoa-Maysuyama, anisotropic diffusion or bilateral filtering are some of the techniques to achieve this.

Another important pre-processing method that is useful for later image segmentation is **edge detection**, to do this one can use a Canny, a Sobel (Figure 2.9b) or Laplacian operator (Figure 2.9c), or even active contours.

2.3.3.3 Feature Extraction

Feature extraction is the extraction of values from images with the objective of reducing the initial data to a set of non-redundant characteristics that can be used for training algorithms. Some features can be found by edge, corner, blob or ridge detection.

An alternative for feature extraction are the local invariant feature detectors and descriptors, that try to identify patterns that differ from the neighborhood [TM08] which more than detecting part of an object, finds keypoints and organizes them according to the following premises [LRdSM16]:

1. Local interest points, keypoints, are extracted in independent manner from both the test and the reference image, and then characterized utilizing invariant descriptors;

2. Invariant descriptors are organized according to each other;

3. Following different procedures depending on the algorithm, the matched features go trough geometric verifications.

Examples of this approach are the SIFT (Scale-Invariant Feature Transform) and SURF (Speeded Up Robust Features) Operators and the Harris and Hessian Detector [GL11].

2.3.3.4 Detection/Segmentation

Detection and segmentation refer to different but intersected aspects of image processing. While detection usually refers to the identification of elements in a image inside object or non-object related bounding boxes [Har15], segmentation consists in the separation of an image into several partitions in that each subset consists on a relevant part of the image [PZZ13].

Existing methods for detection of objects can be grouped into three categories [COR+16]:

- Segmentation –> Detection: the algorithm learns to classify proposed object bounding boxes and then refines their spatial locations. E.g. R-CNN utilizes a Convolutional Neural Network (CNN) to find object proposals utilizing log loss and then uses a Support Vector Machine on the CNN's features where it replaces the softmax classifier in object detection. [Gir15]

- Detection –> Segmentation: CNNs or non-parametric methods [CLY15] start by assigning a category label to the pixels of the image and then uses the bounding boxes.

- Detection *and* Segmentation: traditional approach used in the Hough voting [MC15], more recent implementations use this method by training a CNN [TLFT11] with the objective of identifying more difficult to detect or to segment object classes [HAGM14].

2.3.3.5 Recognition

Although a system can identify lines and parts of an image it remains a challenge for it to comprehend a scene and to recognize all the objects (e.g. differentiating a cat from a deer). It is possible to analyze the recognition problem in three perspectives [SS01]:

- Object Detection (to quickly assess whether there exists a match with an object in an image)

- Instance Recognition (to assess the existence of a specific solid object, with characteristic feature keypoints that can be verified if their positions are in a geometrical acceptable way)

- Class Recognition (for categories of diverse objects, like people and animals)

Considering the scope of the problem presented, we will go through the class recognition problem whose solution starts by the indexation of local features, after their matching, so that it is possible to generate visual words (using an inverted file index). These visual words can be organized using k-means into a bag of visual words, a histogram of the number of occurrences of particular image patterns [CDF+04], to which we can apply concepts such as term frequency by weighting the different words that compose it. Similar approaches have also been applied to text mining problems, so as to extract and classify topics on on-line social networks, such as Twitter [PPS+17, PPSR17].

2.3.3.6 Classification

The vocabulary created previously can be used to assign a meaning to an assemble of recognized labels. For this there must be a training dataset that the system can use to create relations between a certain label and a set of words. After training, it should be possible to test a certain image and obtain better or worse results according to the algorithms used. The **visual semantics** of an image are the result of the information extracted, the semantic pattern models, which can be used to classify it according to the desired objective.

Classification can be supervised or unsupervised. On supervised classification, you train your classifier on a previously classified dataset and a known goal. With unsupervised classification, on the other hand, there is no knowledge on the classification of the dataset, nor the objectives of its classification. Another interesting concept is reinforcement learning, when there is knowledge on the goal but no previously classified data sample. Popular algorithms used are Bayesian classifiers [DGL96], k-NN [CH67], SVM [CV95], Decision Tree [RM07], Neural Network [WRL94], among others. Unsupervised clustering with categorical labels uses algorithms like mixture models, K-means clustering, hierarchical clustering, fuzzy k-means and a mixture of Gaussians [TP12]. This unsupervised learning can be used to divide images according to similar RGB-colours for example, but in the particular scope of this dissertation it might be possible to apply an unsupervised learning data clustering technique to categorize and find patterns in the features previously detected with the classification algorithms. Examples of some of the aforementioned techniques applied to the recognition and classification of traffic conditions, for instance, can be found elsewhere [LRB09, LKR$^+$16]

For validation of results from a dataset, Cross Validation can be used, this method is used for example in prediction problems where there is a training dataset, meaning a dataset data that is known where the training is done, and a testing dataset where the model is tested.

It is also interesting to note there are different open source tools available that can be used for these kinds of classification [WP16]: OpenCV, WEKA and Rapidminer are amongst the most popular ones.

Another aspect to be taken into account is the growth in Deep Learning methodologies to recognize and classify images with new algorithms and improvement of techniques appearing every year [ZGFD17]. Taking into account the recent positive results of Deep Learning in those tasks a survey was made on the main tools used: TensorFlow, Theano, Caffe and Torch.

2.4 Visual Semantics to Assess Urban Indicators

After training the algorithms it is possible to classify the images according to a certain class or several different classes of indicators using their visual semantics. There is a certain amount of related work that can serve as inspiration, from the detection of similarities between images [DF11] to automated image description generation [KFF17].

Table 2.8: Urban indicators inference by Human resources using GSV imagery

Urban Indicator (labeling)	**Reference**	**DataSet**	**Objective**	**Results**
Litter, empty alcohol bottles, graffiti, burned out buildings, abandoned buildings, abandoned cars, and poor-building maintenance	[QMS+16]	GSV imagery of 532 block faces in New York to be labeled.	Evaluate Physical Disorder using Google Street View	Possible to infer on disorder. Temporary indicators were not useful, like litter.
36 items related to walking, cycling, public transport, aesthetics, land use mix, grocery stores, food outlets and recreational facility-related items	[FCR+16]	GSV imagery from 59 neighbourhoods of five European urban zones	Audit of environmental obesogenic characteristics	Indicators entered in conflict (e.g. Green Areas (+ healthy) and Food being sold on street (- healthy) = inconclusive

2.4.1 Visual Urban Indicators

First stage is defining the indicators. As it can be seen in Table 2.8, there is a broad group of different indicators that can help people classify an urban environment. Examples of this are the classification of an environment as disordered, or taking elations on how certain objects are possible indicators of obesogenic characteristics.

2.4.1.1 Additional Indicators

This section analyses Visual Urban indicators that, while still being interesting to review, could not be included previously for both not being a computer vision example and for lacking transversality to be extracted from google street view (GSV).

- **Sustainable Structural Design** - This indicator is mainly concerned with Green buildings and sustainable infrastructures, with a special interest on the origins of the construction materials [PC16]. But considering the architectural variations related to historical reasons, it also raises the interesting point of how a structure can reflect the growth of a city through times. This indicator has potential, but for now it lacks a comprehensive dataset with labels provided by experts.

- **Nightlights** - Mentioned as a way to obtain information on the existence of slums in certain areas [KSvM+17] using night light from images obtained by the International Space Station (ISS). This indicator lacked a broader perspective with contextual knowledge of specific regions. It presents an interesting approach even if it cannot be applied to the mainly daytime images from GSV.

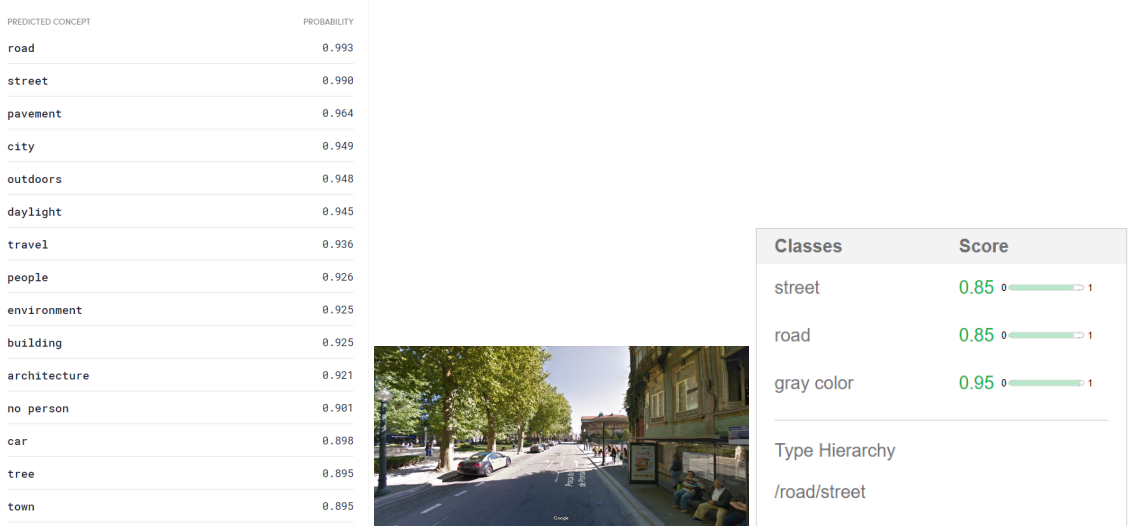
- **Entropy Index** - Although satellite data [KBS15] is for now better suited to monitor and comprehend urban spacial growth, it might be possible to infer on the dynamics of change in a certain area and indicate possible development trends by using an **entropy index** [CSCC16] and applying this methodology to images from the same area in different years.

Table 2.9: Urban indicators inference by Computer Vision from both GSV and non-GSV imagery.

Urban Indicator (label)	Reference	DataSet	Method	Objective	Results	Applicable to GSV sourced Dataset
Sewers	[BPT94]	CCTV survey libraries	Canny Edge Detection	Retrieval of information for non-man-entry (NME) sewer renovation	Possible to identify some variations	No
Sparse vegetation, low-medium vegetation and medium-high vegetation	[Tri17]	Pleaides-1A imagery used	ISODATA Unsupervised And MLEN Supervised	classify vegetation types and estimate vegetation cover percentage in the green zone	Possible to classify Vegetation and classify Vegetation Cover	No
Road Cracking and Surface Deformation	[YWT17]	road video images collected by the camera	Corner detection Harris and SV	Usage of binocular vision	No results included	No
Perceived Naturalness based on a seven point Likert scale from 'very manmade' to 'very natural'	[HRA+17]	Automatically sampled GSV imagery and from the data and images used by Berman et al [BHK+14]	GSV API	high level semantics of objects found in an image have a bearing on PN and therefore possibly on restorative value	Establish the principle that automated sampling and analysis of ground level image data may be useful in exploring perceived naturalness	Yes
Binary; No Curb Ramp, Object in Path, Sidewalk Ending, Surface problem	[HLF13]	GSV imagery + CrowdSource Labelling (dedicated and untrained)	Support Vector Machine	Identify Street-level Accessibility	A mixed CrowdSource works + Computer Vision Works	Yes
"Which place looks safer?"	[NPRH14]	GSV images randomly chosen from the cities of New York, Boston, Linz and Salzburg	Support Vector Regression	Predict Perceived Safety	Results Too Biased	Yes

2.4.2 Application Examples

In Table 2.9, we can already see computer vision having an important role. The first three cases go through pre-processing and then are classified according to the aim of each study. In the last three, the approaches are more similar to what is intended in this project.



(a) Clarifai (b) Image to be classified. (c) IBM Vision Recognition API.

Figure 2.10: Image classification by IBM and Clarifai.

The most important aspects of these projects are the calculation of Perceived Naturalness, that is, people's perception of the environment characteristics[OFT+09], and its indicators. For that, the Google Street View API imagery is used to show human voluntary helpers pictures for them to classify, and then elements that are part of the image are extracted in order to establish patterns.

It is important for this phase to have an accurate database. There are several that can be used for an urban environment, like Cityscapes [COR+16], a large-scale dataset to test and train semantic labeling. Another alternative is using GSV to obtain indicators that can be organized into different class labels [HLF13]. Clarifai and IBM Vision Recognition API are others alternatives to obtain indicators, like it can be seen in Figure 2.10.

2.5 Summary

Using the systematic literature review process to organize the literature review research created an opportunity to find some studies that otherwise would remain unidentified. A weakness felt is that the choice of keywords might not have been the best, since it returned a too high a number of results, making it impossible to filter all of them within the time frame. Not only that, but the results, while being good to list indicators, failed in providing the concepts and methodologies useful when understanding computer vision and its applications. This might threaten the research, especially when considering the creation of a list of visual urban indicators accurately classified in terms of importance. Another aspect that might have impacted the results was not limiting the sources for papers. Perhaps a different approach in the future will limit the sources to a smaller number.

In relation to the second section, it is possible to observe that different kind of indicators, for example of sustainability, will probably be used to establish semantic classes for visual urban indicators. Several areas of interest were identified being mostly rooted in Environment, Society and Economy.
The second section explored the evolution of the study of computer vision thus allowing a general grasp of the change of approaches facilitating the understanding of the methodology to extract visual semantics from an image, being that it lacks a much more extensive study of the deep learning algorithms that will probably have to be explored in the implementation of the proposed approach to the problem.
Finally in the third part it was possible to see related work to what will be attempted in the future. In addition, supplementary urban indicators were analyzed with respect to their applicability to being inferred using computer vision from Google Street View imagery.

In conclusion, it is possible to make the following assumptions for the future methodology:

- It is possible to organize visual urban indicators in a taxonomy based on sustainability indicators;
- Easy access to datasets, but need of adaptation to the particular case, extraction of images from GSV shows no notable issues;

- Although less explored in this context deep learning algorithms appear to show the best potential for a good performance.

Chapter 3

Methodological Approach

The aim of this chapter is to provide an extensive characterization on the outline of the project, and identify the methods used in it. From the data collection to the demonstration of results, by the end of the chapter it should be possible for the reader to understand the general scheme of the final result.

Although this chapter is dedicated to the methodological approach, some inferences are made on the requirements of the usage of Google Street View, since the technology is intrinsic to the scope of the dissertation.

3.1 Proposed Architecture

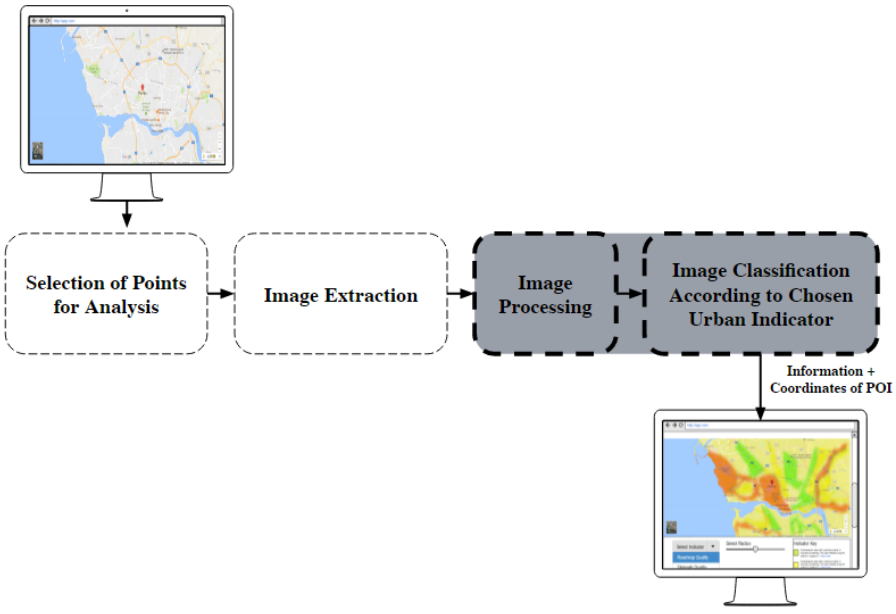


Figure 3.1: Preliminary Methodological Approach

The proposed methodological approach starts with the selection of an easily identifiable urban indicator. A good example of such is the existence of green areas in a city, which is a great indicator of sustainability [GJM+11] and of better quality of living for the habitants of that area [MVdV+09].

The following step in the approach, as can be seen in Figure 3.1, is the selection of points for analysis, from which the images will be extracted.

We then can infer urban indicators from these images and process and classify them accordingly.

Finally, the software returns feedback to the user, using the newly obtained image classifications, along with the coordinates of the point of analysis.

The previously mentioned steps were found to be incomplete, and did not reflect the possibilities of integration with other methods of independent automatic extraction, besides from visual extraction, nor did it reflect on the series of actions to achieve a correct visualization of the urban indicators. As such, the proposed architecture was extended to go into further detail on these different aspects, as seen in Figure 3.2.

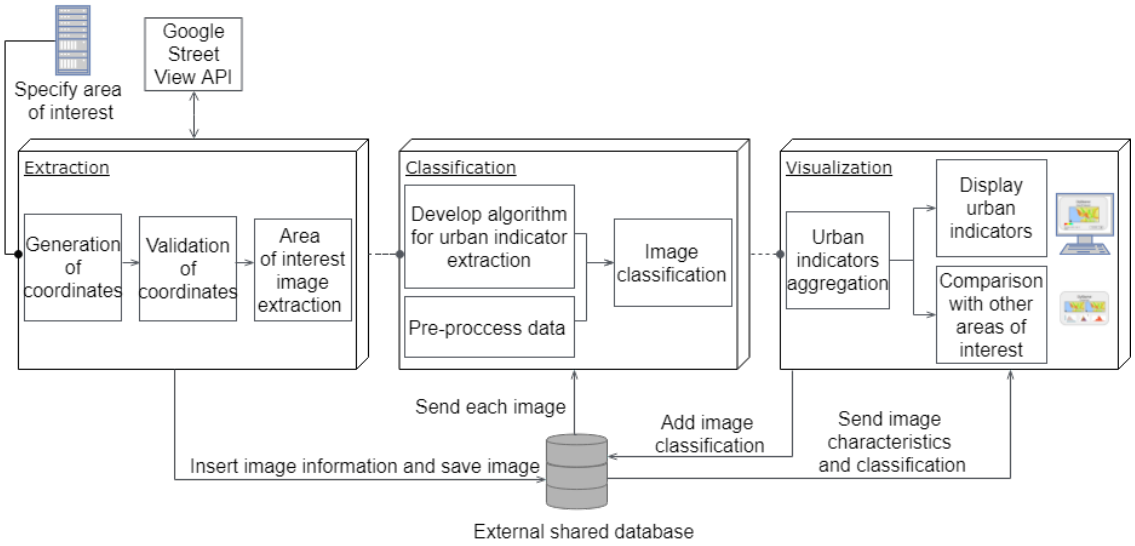


Figure 3.2: Proposed Architecture for Methodological Approach based in 3 modules

The first module is focused on the extraction of data, the basic challenges that are involved in it and how much data is perceived to be necessary for a case study.

The second module is responsible for the classification and should present methods that are effective and efficient, in regards to financial costs as well as necessary computing power.

Finally the third module concludes with the visualization of data. It focuses on how to present information in a way that makes it easier to ascertain the quality of life of a region, taking into account its urban indicators.

Each module should work independently, since although this project is centered, at the moment, around Google Street View data extraction and comprehension, other geolocated data can be used and classified using different methods. This architecture should first be tested with a small sample to assess on its viability.

3.2 Description of Architectural Modules

3.2.1 Extraction

One of the objectives of this dissertation is to study the viability of the image extraction both when faced with a comparatively simple study case, as well as with a more complex one. As such, the first phase of the extraction of the data should be performed with a small group of images to be posteriorly classified and interpreted. Said interpretation should be made from the perspective of the possible urban indicators that can be extracted, so that it is possible to translate it to indexes of the urban characteristics, which in turn are useful for human understanding.

Firstly, when choosing the area to be analyzed, one should take into account the following aspects:

- **Boundaries**, which are well defined limits in latitude and longitude that allow Google Street View to know where to extract images from;
- **Diversity** of the area's characteristics. A large variety of characteristics allows its division into different sections (green areas, buildings, water sources). Lacking this variety, the area to be studied should offer distinct attributes that will serve as a contrast when comparison to other case studies;
- **Availability of other information sources**, that is, if there is a means of validating the results of the data extracted and processed.

In regards to the extraction of the images from Google Street View, it should be taken into account that, before loading the image, one should request for the image metadata. The requests for the metadata provide a free to use mean of checking if the coordinates return a unique non-null image, since Google Street View, besides being limited to areas where there are pavements, has yet to secure full coverage of areas like Northern Asia and Central Africa [Goo17a]). Besides the status of the image, it is also possible to extract the date in a "YYYY-MM" format. This is important as it offers a possibility of filtering old, and possible outdated images, or at the very least to offer more certainty on the deficiencies of the final result.

Finally, both the images extracted as well as the localization of the file and the information taken from the metadata (coordinates and date mainly) should be saved in a database, like it can be seen in Figure 3.3 diagram. Its important that the image is saved in this phase of the project since, while the classification can be done in almost instantaneously when using the right tools to achieve classification, the area classification as a whole needs a great input of images.

3.2.2 Classification

The objective of the classification module is to extract and organize the indicators of the data in a way that they can be used by the visualization module to enable human comprehension. As such it was perceived that this module would be divided into two phases:

- Obtaining a Classification Algorithm & Extracting indicators;
- Automatically organizing the indicators.

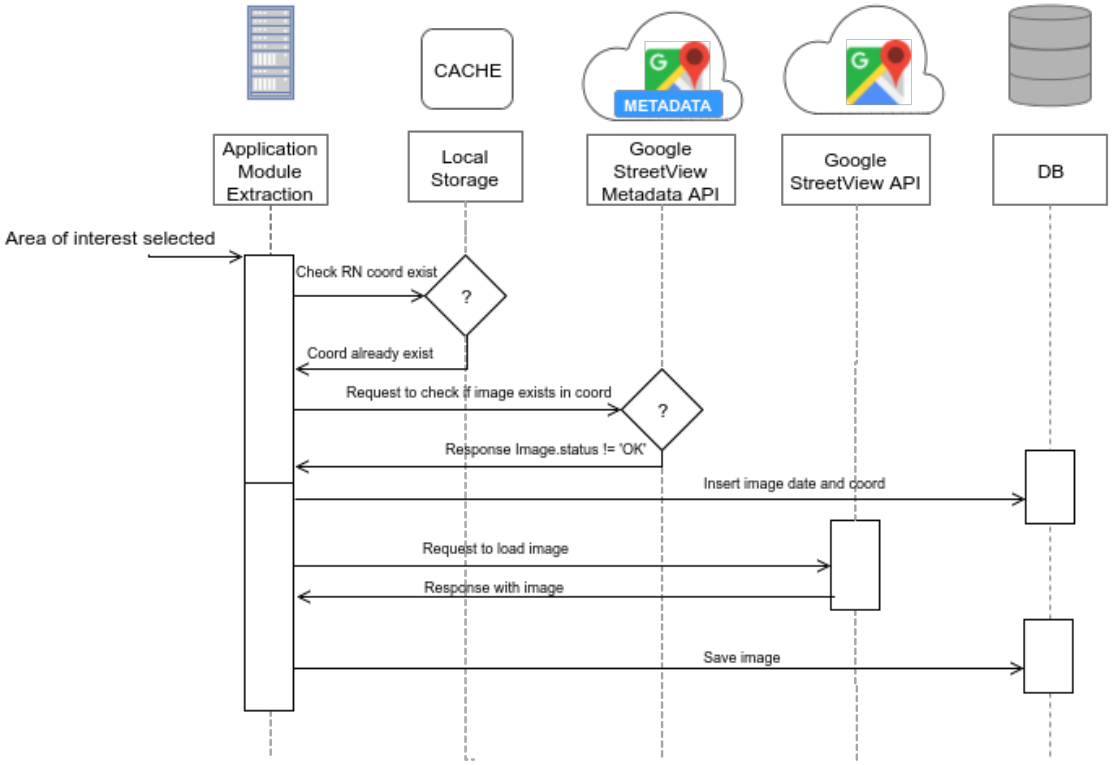


Figure 3.3: Extraction Module Sequence Diagram

3.2.2.1 Classification Algorithm & Extracting indicators from the image dataset

To obtain a suitable classification algorithm, three approaches were considered. The first approach was to train and test a classification algorithm. The main issue present in that approach is that image classification, in order to work in the scope of this dissertation, needs a suitable dataset with multi-tags related to urban characteristics to be appropriately trained, as it can be seen in Figure 3.4, which may be complex to find. The second approach involved choosing exclusively one characteristic and applying image preprocessing techniques, like filtering and thresholding, with the objective of extracting information. This approach would not be transversal and would be limited to one study case. A third possible method would consist not in the development of an algorithm, but in the application of an already existing external API. Considering that the main objective of this phase is to extract indicators,it is an important alternative to be mentioned.

Optimal training dataset

Since this project considers the utilization of Google Street View, it should result in an analysis with some advantages over the more conventional approach using satellite images. An example of a case where this does not happen is the study of the exact size of green areas. In this case, satellite images provide a clear advantage since they can view the whole area and not be limited by images taken from the street, which would provide insufficient data on this regard. Having this in mind, in terms

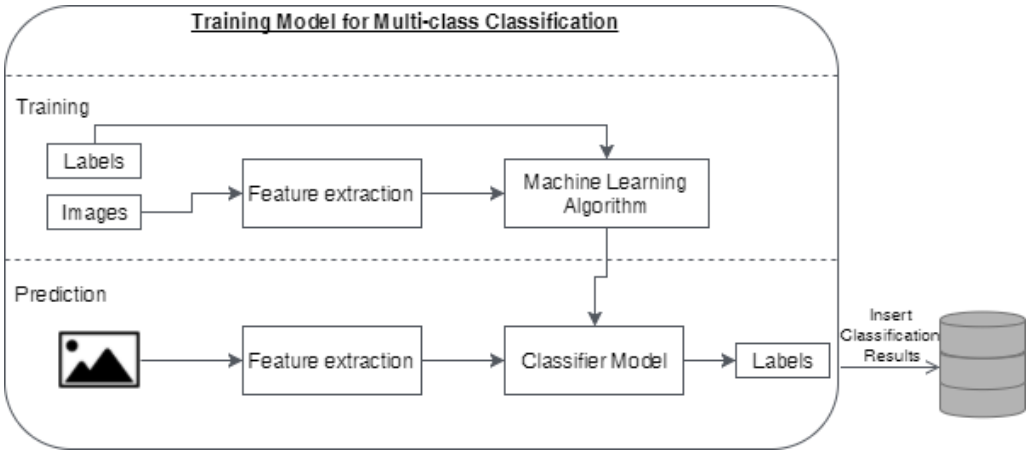


Figure 3.4: Classification Module by using a labeled dataset to train the classification Algorithm

of training dataset, the optimal approach would take advantage of the closer point of view provided by GSV. Thus it would be capable of asserting on the characteristics of the buildings, the degree of degradation of the streets or the number of cars identified in the images.

To be capable of obtaining an acceptable training dataset there are several possible methods:

- Manual classification by experts;
- Preexisting set from street scenes;
- Extracting information from real property/estate sources.

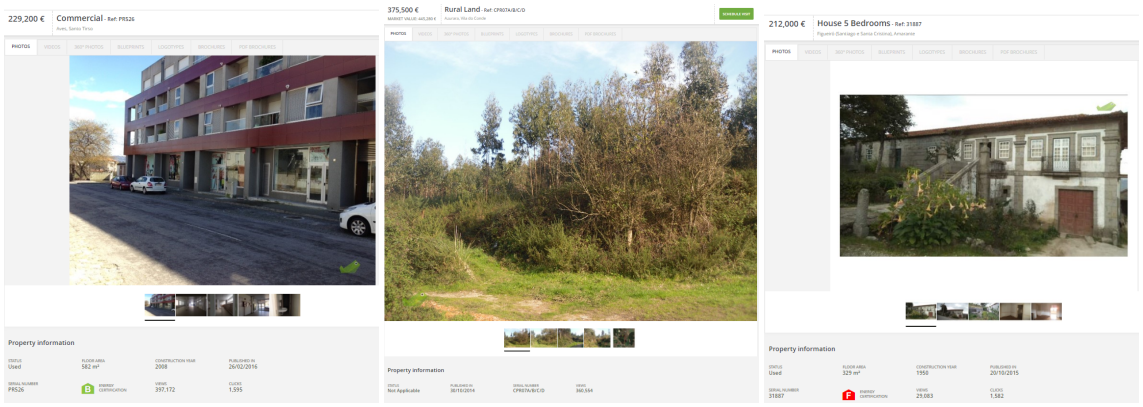
The ideal training data for a supervised learning would have a **manual classification** by an expert in urbanism, for example with an architectural point of view, in order to be capable to infer on the visual cues that indicate the management and physical design of urban structures. This specific data labeling is high resource consuming, since validation of both the possible labels and the specific identification in each image depend on human based resources. However, it has the best potential results, since the data to be classified could be images directly extracted from Google Street View, unlike the other two methods.

Since there is a lot of interest in automotive driving, there are datasets from **street scenes** that focus on the characteristics that can be found on the side roads. In

Table 3.1: Class Definitions from Cityscapes Dataset Overview

Group	Classes
flat	road · sidewalk · parking · rail track
human	person · rider
vehicle	car · truck · bus · on rails · motorcycle · bicycle · caravan · trailer
construction	building · wall · fence · guard rail · bridge · tunnel
object	pole · pole group · traffic sign · traffic light
nature	vegetation · terrain
sky	sky
void	ground · dynamic · static

Figure 3.5: Comparison between images from taken from GSV (left) and Cityscapes(right) dataset. *Berlin, coordinates 52 30 57.9"N 13º22 32.7 E.*



(a) Modern Building (b) Rural Land (c) Old Building

Figure 3.6: Example of data that can be extracted from Real Property website.

particular Cityscapes [COR+16] appears to have images taken from a similar angle to what would be achieved using Google Street View, as can be seen in Figure 3.5. The groups of characteristics and the labels this dataset include, listed in Table 3.1, also seem to be acceptable to be used as urban indicators, since they include both nature and human made structures and it differentiates between them.

For the alternative of a dataset based on **real property sources**, it implies the construction of a crawler that will extract information on market value and general characteristics of a specific property. This approach must make some assumptions based on the general organization of such sources (e.g. websites). In Figures 3.6a, 3.6b and 3.6c, it can be seen examples of the information present in an example of a real property website [1]. It is possible to observe some patterns: the first image is generally the image of the front of the house, each house has a date of construction, a property price and floor area. It is also possible to observe that properties that have "status=Not Applicable" are generally green areas which can also be used as a filtered when preparing the dataset for training. This type of dataset can be used not only for training a classification algorithm, but also to make other observations on urban area characteristics.

[1] https://casa.sapo.pt/en_GB/Imoveis-Banca

Pre-Processing Based Approach

While the previous approach is defined by the dataset, this one is limited to the context of the chosen indicator. Like previously stated, to get the results on the exact size of a green area of a city, images extracted from GSV are not optimal, since although consisting on omni-directional imagery, they do not offer an image distanced enough to be possible to get completely accurate results. Still, the influence of the perceived presence or lack of green spaces in the psychological state of an individual can be used as an indicator. As such, techniques with small complexity based on thresholding to exclude pixels not of interest can be applied and posteriorly worked with. In the case in point, the percentages of green area present in each image would be an interesting characteristic to analyze. To be noted that this approach is far from ideal, but the implementation is straightforward and does not require much computational power.

External Image Recognition Application Programming Interfaces

Unlike the previous approach, the current approach requires the usage of an external API. The interface should be able to identify the main characteristics of an image. However, since indicators could also be related to less distinguishing characteristics, like the color of the background or if it's cluttered, returning more tags even if less accurate can still be useful for the subsequent organization of data. This is particularly noteworthy when taking into account that, usually, image recognition APIs are used to organize images according to the general classes presented in them. So, if a dog appears in a central focus of the picture, it is more likely that the API will concentrate on the tag "dog", "animal" or even the "breed", which for the scope of this dissertation is not very useful. This method does not require much computational power, since the training of the model is done externally.

However, it is often limited by quotas and possible budget restraints. A comparison of results between some of the current main image recognition applications in the market can be seen in Table 3.2, where three images were chosen as examples of the type of images that will be classified. It is possible to see that one shows a road with cars, the second a wall with trees and finally a third with a simple 2-story building, showing its door and windows.

Between the four chosen APIs to be compared, we can conclude that CloudSight has results too specific and descriptive to be directly used in a classification in which images will be grouped according to the feature they present. IBM Watson Visual Recognition was capable of describing colors but provided a very specific class identification was not always useful: for example, 4 out of the 11 labels were variations of hotel terms. This tool also failed to identify more varieties of nature present in the second image, and not capable of identifying the cars at all in the third one, although it gave a very accurate description of the main building. The Google Cloud Vision API provided the most interesting tags, even attempting to find the types of the cars present in the image. This diversity of choices could be helpful in identifying transportations system related classes. Clarifai had good tags (was able to identify the presence of the main objects of each images, and also the more relatively minor objects like the grass in the second image), albeit not so conclusive when compared

to Cloud Vision. Additionally, it also misidentified water in the second image, possibly because of the distortion. Even so, the less specific labels might present an advantage, by being more transversal across different cultures and environments.

An additional aspect to be compared is that IBM Watson Visual Recognition has the best free usage quota for month but the requests are limited per day which considering the results might not justify the value (CloudSight - 500p/month; Google Cloud Vision API - 1,000p/month;Clarifai - 5,000p/month;IBM Watson Visual Recognition 250p/day). Another aspect to be noted is that the Clarifai API offers the opportunity to train a personalized model. These characteristics, among with the versatility of its labels, make Clarifai the most suitable tool from the ones considered, for an exploratory proof of concept.

3.2.2.2 Automatically organizing the Indicators

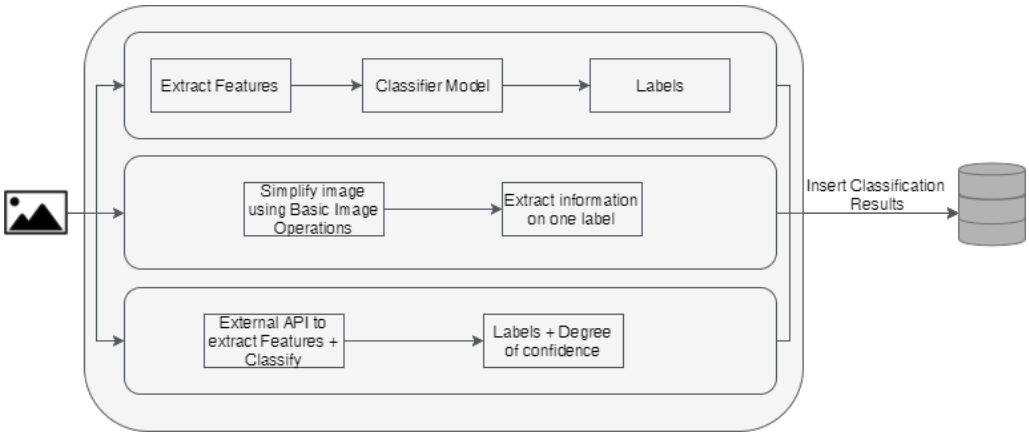


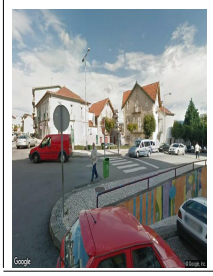
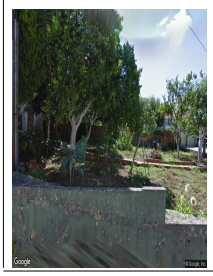
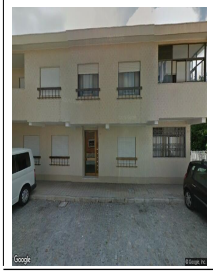
Figure 3.7: Classification Module taking into account different Approaches

For portability purposes, the results of each approach should be stored in a table similar to the one resulting from the image extraction, with the date and the localization of each image. This also enables the exploration of patterns that allow for the data to be clustered, which would contribute to the visualization of the final results. Ideally, this division should be made taking into account the different types of urban environments. However, this would require a previous classification by specialists. Considering the inability to acquire such labeling, a different approach could be made to find a structure in the results, taking advantage of unsupervised learning techniques such as k-means clustering [JMF+99]. This algorithm can be used on unlabeled data to obtain a selection of objects akin to each other in an attempt to find patterns.

3.2.3 Visualization

After the classification of the elements present in each image, there is a possibility that the elements by themselves might not produce a suitable urban indicator. Given

Table 3.2: Comparison between different Image Recognition APIs

	CloudSight	Google Cloud Vision API	Clarifai	IBM Watson Visual Recognition
	"hatchback red of 5 doors"	Car Property Family Car Vehicle Transport Town Residential Area Luxury Vehicle Neighbourhood Sky Road City Car Suburb Asphalt Real Estate Compact Car Sedan Subcompact Car Roof Road Trip Minivan	car outdoors horizontal plane transportation system road street travel architecture luxury house traffic pavement city sky tourism no person stationary vehicle asphalt town	Parking Lot dwelling housing motor hotel hotel building motel Hotel Building vehicle reddish orange color steel blue color
	"tree with green leaves"	Property Tree Vegetation Plant Yard Residential Area Landscape Land Lot Real Estate Area Grass House Home Backyard Outdoor Structure Village Plant Community Landscaping Cottage	nature water tree outdoors summer wood calamity house grass garden leaf flora rain environment travel architecture tropical demolition flood	Storm nature Yard dwelling housing retaining wall wall fence barrier gray color
	"gray concrete structure"	Property Luxury Vehicle Building Family Car Home House Residential Area Car Real Estate Facade Sedan Area Commercial Building Window Estate Mid Size Car Elevation Executive Car Vehicle Villa	architecture outdoors home luxury parking lot indoors car garage absence door window abandoned house empty horizontal plane contemporary offense wealth family	garage building ground floor floor duplex house beige color

this, in the visualization module there might be a need to aggregate the characteristics of the area into sections capable of being displayed. This aggregation might take advantage of the clustering techniques previously mentioned.

This dissertation aims to achieve the visualization of data with recourse to two different approaches. One is by creating a user friendly platform, centered on the choice of an urban characteristic and view of its distribution in a certain area, by processing images extracted from Google Street View and then classifying them. The other is by comparing the urban indicators profile of an area with another to find similarities in their characteristics. This should be possible in both distanced areas as well as neighboring ones for which obtaining the administrative division, like district distribution, is crucial. This analysis is particularly interesting since one of the problems of the extraction of urban indicators is the lack of standard quantifiers that allow for a straightforward comparison.

3.3 Summary

The objective of this chapter was to present the methodology proposed for creating a system capable of extracting urban indicators of an area using Goggle Street View. As aforementioned in this chapter, the methodology proposed for this project is currently defined and seems to be versatile and promising. As such the workflow for each of the proposed modules was idealized in a way that it would currently work independently from the others, which is particularly important considering that the components for developing a model for classification are greatly dependent on the pre-existence of a suitable dataset with high-level features correctly tagged. Bearing that in mind, for the training of the classification models three different extraction methods were collected; however, from the methods considered, only the extraction of images through crawling a website was effectively automated. The other, nonetheless, pose important challanges for future developments.

Chapter 4

Implementation, Results and Analysis

This chapter addresses the implementation of the project considering its limitations and results. It follows the structure previously presented in the methodological approach. For this reason, it starts by explaining the image extraction process, the classification and visualization tools and their specific settings. It then presents the results of the current implementation, at the time of writing. The discussion then goes into the interpretation of the obtained results, along with commentary on their quality and limitations. In addition, different approaches that were attempted but did not offer satisfactory results are also mentioned in this chapter. An example of such is the training of the algorithm using a dataset extracted from a real estate website.

4.1 Current Implementation

The implementation is centered on each module achieving the goal defined in the architecture proposed in Chapter 3. The general technologies applied to each module can be seen in Figure 4.1, for the sake of illustration. Their specific role will be explained in the following section in more details.

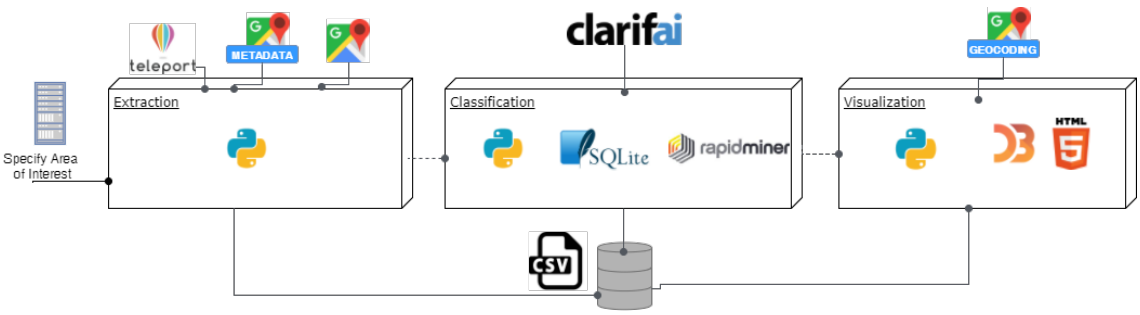


Figure 4.1: Technologies of Current Implementation

Extraction

In relation to the extraction module, the system was mainly implemented in the Python[1] language. This was a project decision due to vast number and diversity of state-of-the-art machine learning liberaries available in Python. In addition to that, Python is easy to use, is supported by good documentation, and count on an enourmously engaged community [PVG+01].

The geographic coordinates of the urban area were automatically obtained by making use of Teleport[2], a location information online database. As previously stated, this dissertation analyses urban indicators retrieved from Google Street View imagery. For this effect, the extraction module also made use of the Google Street View Image API[3] and the Google Street View Metadata API[4] to obtain information about the images in the defined coordinates, and to adequately extract them.

The images for the dataset were extracted in a 640x480 resolution in order to offer good visualization quality, without using too much space. For testing purposes, the sample size was 100 images, although a subsequent sample of 10,000 images of the general area of interest was also taken. The images were locally stored and the information on the latitude, longitude, month and year was saved in a comma separated values (CSV) file.

Classification

In order to carry out the classification, the extraction of high level features was made using the Clarifai General Model[5], which is a classification model capable of identifying 11,000 different concepts. A concept, in this context, is a class associated with an image (other names for this notion are tag or label). Considering that it was decided that having a greater number of elements identified in an image, used to later find patterns between different scenes, was more desirable than having an extensive description of the main object of the scene. Such a large amount of classes offered an advantage when compared to other applications for external image recognition. The usage of Clarifai allowed us to bypass a potentially time consuming training of a machine learning algorithm from the ground up. The output is a light-weighted JavaScript Object Notation (JSON) file with the name of each concept (or *label*) identified in the image, and its degree of certainty (0 to 1). The general structure of the file can be seen in the listing below:

```
1 {
2   "status": {...},
3   "outputs": [ ...
4       "data": {
5         "concepts": [
6           {
```

[1] https://www.python.org/
[2] https://developers.teleport.org/
[3] https://developers.google.com/maps/documentation/streetview/
[4] https://developers.google.com/maps/documentation/streetview/metadata
[5] https://www.clarifai.com/

```
            "app_id": "main",
            "id": "ai_Zmhsv0Ch",
            "value": 0.9931723,
            "name": "outdoors"
          }
      ]
      }
  ]
}
```


The information previously extracted from the images was loaded from the CSV file onto an Sqlite[6] database, where it was complemented with the classification results. A new CSV file was then generated from this process.

Considering that the Clarifai general model returns a great number of concepts for each image, arranging them into different groups would be useful for future analysis. Although such clustering could be manually processed, and considering that the purpose of this dissertation is automating the urban indicators extraction and thus sparing some of the manual labor, the approach of using an unsupervised machine learning method was preferred. Specifically, the k-means cluster algorithm was chosen. For this purpose, the open-source data science software Rapidminer[7] was used, since it has a good implementation of the k-means algorithm and allows for an intuitive workflow, simple to modify and experiment on.

The workflow seen in Figure 4.3 consists of three parts: i) the processing of the data taken from the CSV file with the high level features; ii) the creation of the clusters; and iii) applying those clusters to the dataset to finally sotore in a database.

The processing of data for the clustering algorithm consisted in selecting the label columns and dealing with missing elements. This last step is particularly relevant considering that the k-means algorithm expects only numeric values. Since missing elements in this table are concepts the Clarifai model did not consider relevant, and it is normal for each image to lack a large number of the labels (from the 11,000 total available in the case of the Clarifai general model), techniques to deal with missing values were applied. Techniques such as replacing missing values with the median value of other examples, or removing the objects where the values were missing, would result in a complex manipulation of data. The manipulation of data would occur because Clarifai returns the first 20 labels only and provides no data on the other remaining labels in the dataset. For this reason, removing the objects in which there are missing values would delete the full dataset. On the other hand, replacing missing values with the median value would not differentiate between labels that are not present at all in the image from those effectively there, but not appearing among the first 20 ones returned by Clarifai. To avoid that, the approach taken was to consider missing values equal to zero whenever there was a lack of numeric values in the dataset.

[6] https://www.sqlite.org/
[7] https://rapidminer.com/

The clustering operator receives this processed information with only the labels and their confidence values and runs the algorithm. The parameters for the clustering, for testing purposes, were a k value varying from k=2 to k=10 and maximal number of runs of 60, making use of random initialization.

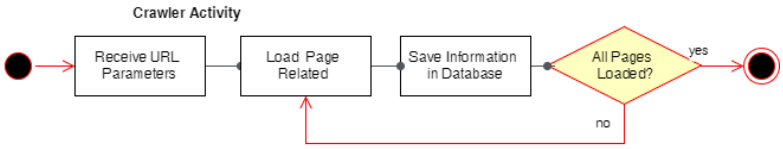


Figure 4.2: Crawler activity diagram to extract alternative training dataset

Alternative training dataset retrieval

With the objective of proposing other methods to extract interesting information on urban characteristics from the Google Street View imagery dataset, a possible classification training dataset was extracted using the crawler technique on a real estate website. Although this alternative dataset was not used for the training of the classification algorithm, and as such should not be considered part of the modules for the current implementation workflow of the project, this approach is to be used in future iterations.

The idea of creating a dataset by taking advantage of the official classification in real estate websites seems to be very promising, as it is expcted to allow for the classification of the houses of a section of a city according to their characteristics. As such, this kind of website was chosen, since its information is deemed to be accurate and provides a general view of the building that can be used to extract labels for training an algorithm. An example of this would be to correlate the year of construction to the architecture of the facade of a house. Another example, and perhaps more interesting in the scope of this dissertation, would be to see how the urban indicators of an area affect the general prices in the real estate market.

The creation of this dataset was achieved by developing a crawler that goes to a real estate website[8] and looks for information on the properties for sale, extracting it. The structure of the crawler can be seen in Figure 4.2.

This script was done in Python taking advantage of its library Beautiful Soup[9], for pulling data out of HTML. The information that it returns is the main image, which is the facade of the building taken from the street point of view in most cases, the price of the property at sale and the specific URL of the property details. For future installments, the URL property details page will return other characteristics such as the year of building and the area size, to name a few.

Visualization

The visualization step consisted in the visual presentation of the information from the database. This was done in two fronts, as described below.

[8] https://casa.sapo.pt/
[9] https://www.crummy.com/software/BeautifulSoup/

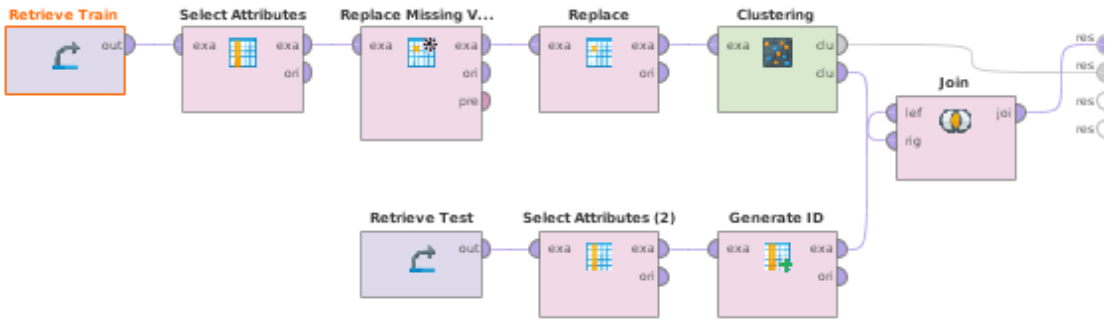


Figure 4.3: K-means Process in Rapidminer

The first one is a website created using mainly HTML5 and a JavaScript library, D3.js (Data-Driven Documents)[10], for the manipulation of documents based on data. This visually presents the variations of a given indicator in a map, providing the user with a general idea of the main characteristics of an area.

The other way of visually interpreting the inferred urban indicator information on a defined area is first by obtaining the administrative divisions of the area under consideration. Then, according to the results of the clustering algorithm, display tables and data charts with the information pertaining to these divisions. To do this, there was a need to extract from the coordinates information such as the locality, district and parish of each point of interest. The results were achieved by creating a Python script that used the Google Street View geocoding API[11].

4.2 Results and Discussion

The main results of the implementation of the proposed approach are detailed in this section. In addition, this chapter also discusses possible further developments that can follow from the current results.

4.2.1 Extraction Results

The area selected for the sample was the city of Porto, Portugal. The city has a diverse mixture of water sources, green areas, old buildings and modern ones near the city center, making it a suitable example for this analysis. The area delimitation boundaries for the city, retrieved from the Teleport API, can be seen in Table 4.1. East and West are the extreme latitude values, while North and South represent longitude limits of the bounding box.

As shown in Figure 4.4, the actual extracted images fall also outside the region of the district of Porto. This happened due to the simplification of the actual region boundary of the district and considering instead a rectangle area containing it, for

[10] https://d3js.org/
[11] https://developers.google.com/maps/documentation/geocoding/

Table 4.1: Coordinate Boundaries of the Porto Region (Sample A)

East	North	South	West
-8.212747	41.392419	40.823076	-8.778112

ease of integration with the Teleport API. However, through identification of the district of each point, it is possible to filter out the points not belonging to the district of Porto, and later use them for a comparison between areas. The choice of the coordinates of the images to extract was made at random. The first sample consisted of 100 unique points (Figure 4.4b) of total size 4.45 MB, and then a larger sample of 10,000 points (Figure 4.4c), occupying a size of 536 MB was also collected. The distribution of the years when the images were taken can be seen in Table 4.3. No filtering was done taking into account the age of the images since it could exclude important areas that have not been updated.

To be noted that, due to quota limits, most of the experimentations in the following classification and visualization modules were made in Sample A of 100 images. The full 10,000 imagery dataset from Sample B was not completely used, but instead a smaller sample of 1,055 (B_1) images was considered for more advanced training and testing purposes. A third sample (Sample C), of a different geographical area, was later included to compare some aspects with Sample A. This Sample C consisted of a region near Istanbul, and its bounding box coordinates can be seen in Table 4.2.

Table 4.2: Coordinate Boundaries of the Istanbul Region (Sample C)

East	North	South	West
29.678	41.42	40.738	28.313

Table 4.3: Distribution of Images Extracted according to Years

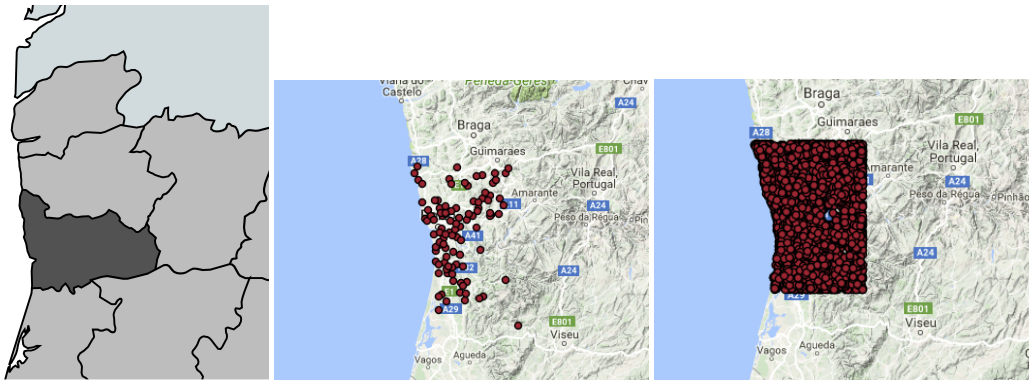
	2009	2010	2011	2012	2013	2014	2015	2016	2017	**Total**
Sample A	31	20	0	0	0	34	11	2	2	**100**
Sample B	2460	2761	0	0	0	3885	834	16	44	**10000**
Sample B_1	270	290	0	0	0	409	81	1	4	**1055**
Sample C	0	0	0	0	0	72	26	2	0	**100**

4.2.2 Classification Results

Like previously stated, besides the classification and extraction of all visible high level features, to identify which would work as urban indicators, it is also necessary to analyze the meaning of the retrieved concepts as well as their distribution. As such, in the current implementation, the results can be divided into:

- Extraction of indicators and its direct study;
- Clustering of the indicators to automatically develop a useful division.

For the sample of 100 images in Sample A, the Clarifai tool returned a total of 178 labels, for which the complete list can be seen in Table 4.5. Those labels can be related



(a) Porto District Area (b) 100 image sample (c) 10000 image sample

Figure 4.4: Geographic distribution of extracted images samples in comparison to the Porto District Area.

to indicators: for example, by ascertaining the absence or presence of labels such as “car”, “airport” or “traffic” we can infer an area’s access to a transportation system. To be noted that Clarifai only presents the 20 tags with the highest probability (or certainty) scores, on the likelihood that an image contains each label. The minimum threshold is not predefined, so a label could be considered, yet still have a low value of probability of its presence in the image. However, it was observed that the probability of the presence of each label in Sample A varies between 0.7611727 and 0.999078. In this particular case, it means that once identified in an image, the certainty of a label being present, and the classification being accurate, is always superior to 76%. To analyze the scope of results in this area, a transformation was done by converting to 1 all values higher than 0. It is possible to view in Figure 4.5 that, in the 100 sample dataset, there are some values that appear with a great predominance. By analyzing the data, it is possible to conclude that the predominant indicators in this sample, appearing in more than half the records, are the ones listed in Table 4.4. In this table, it is possible to also observe to the nearest 3 decimal digits the minimum and maximum value of certainty of presence of that label in each image where it is identified.

Table 4.4: Top Concepts Name Occurrence and Range Interval for Sample A

Name	**Occurrence (%)**	**Value Interval**
outdoors	96%	[0.817 - 0.989]
no person	93%	[0.833 - 0.997]
architecture	90%	[0.842 - 0.998]
house	89%	[0.842 - 0.998]
building	79%	[0.809 - 0.988]
travel	75%	[0.834 - 0.990]
street	64%	[0.808 - 0.991]
family	61%	[0.807 - 0.990]
home	59%	[0.809 - 0.996]

To be noted that having these labels appearing for almost every image may make them not very relevant to study and categorize the area. However, it is likely for them to have a lower frequency when analyzing a different area, making the labels

Table 4.5: Full list of labels returned from classification of sample A

outdoors	horizontal plane	architecture	street	city	travel	transportation system	road	car
wood	door	empty	indoors	dirty	design	expression	industry	abandoned
wealth	parking lot	patio	balcony	wear	room	desktop	texture	furniture
growth	safety	museum	warehouse	police	square	park	driveway	doorway
garden	luxury	stationary	vehicle	asphalt	wall	water	tree	river
bungalow	garage	sunny	brick	winter	truck	reflection	flora	fair weather
expressway	bitumen	pollution	competition	action	danger	war	military	cemetery
business	home	restaurant	food	seashore	beach	trip (journey)	table	rug
yard	lawn	building	modern	apartment	window	construction	estate	front
sky	people	offense	insurance	property	entrance	vertical	highway	commerce
tourism	urban	old	mansion	accident	blacktop	stock	contemporary	absence
no person	storm	demolition	roof	summer	bus	sea	guidance	security
flood	weather	family	leaf	rain	tropical	commuter	train	station
bar	hotel	resort	chair	seat	facade	grass	signal	hurricane
villa	traditional	clean	residence	motion	concrete	environment	calamity	suburb
gutter	glass items	soccer	sunblind	palm	lake	agriculture	daylight	exterior
auto racing	countryside	house	pavement	pattern	picture frame	sight	footpath	ancient
fence	horizontal	bed	tile	interior design	blank	nature	landscape	partition
religion	battle	real	shopping	shop	election	stone	rural	town
residential	airport	light	shadow	vacation	gate	traffic		

more useful for that analysis. To check this, the main labels of samples B_1 and C were also extracted, and listed in Table 4.6. Although Sample B_1 is from the same region as Sample A, only with more points of interest extracted, it has differences in the top labels, thus laking, for example, the labels “building”, “street”, “family” and “home”. On the other hand, the region of Sample C only lacks the term “family” from its list, when comparing to Sample A. In addition, both Sample B_1 and Sample C have an additional label of “sky”. Another interesting observation, is the prevalence of Nature-related labels in Sample B_1 such as: “sky”,“tree”, “nature”, “landscape”, “summer” and “grass”. The similarity between sample A and C suggests that sample C might be organized into similar clusters, despite being different and distant regions.

Table 4.6: Top Concepts Name Occurrence for Sample B_1 and C

Sample B_1		**Sample C**	
Name	**Occurrence (%)**	**Name**	**Occurrence (%)**
outdoors	98%	outdoors	98%
no person	97%	no person	97%
travel	87%	travel	89%
sky	75%	architecture	65%
tree	70%	house	59%
nature	65%	road	57%
landscape	64%	building	57%
summer	63%	street	53%
architecture	54%	sky	50%
house	51%		
grass	51%		

Unsupervised Machine Learning on Cluster Indicators

The main objective of using the k-means algorithm in this methodology is to group the images after having extracted the indicators provided by the classification model. For this, a suitable k value had to be chosen in a way that does not result in too much variance inside the same cluster, nor in over-fitting. This last situation would risk

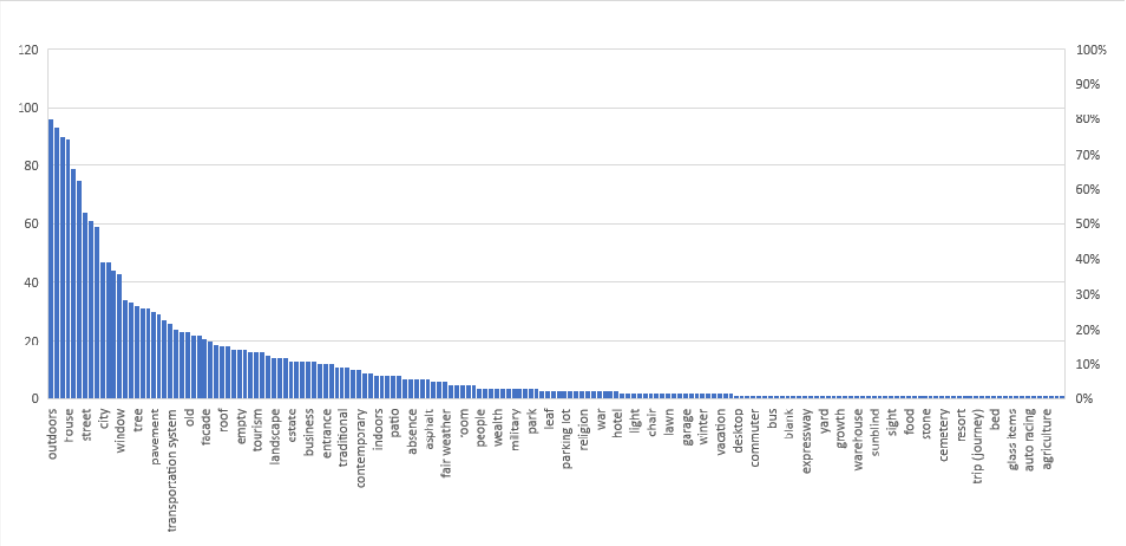


Figure 4.5: Distribution of the labels extracted by Clarifai

making the division too similar to one exact category of data extracted and make it more difficult to predict future divisions.

After considering the issues with choosing a k value too low or too high, one has to analyze the intermediate values. Like previously stated, there is no standard way of choosing an ideal k value and, as such, it is useful to instead make a choice adapted to the scope of this dissertation. The first attempt was by choosing the most characterizing features of the clusters and checking if they varied significantly. By most characterizing features, we understand to be those with a degree of confidence of occurrence superior to 60% in the clusters. Although this approach served to better understand the attributes of the clusters, it was not very useful in deciding the most appropriate k value. Fundamentally, a different method had to be attempted by carefully observing the structure of each cluster.

In Figure 4.6, where k=4, it is possible to see some differentiation between the distinct clusters, in particular between cluster 1 and the others. However, there are some overlaps between them, or peaks near each other, which should be avoided because it implies that the clusters are too alike. Although in a small sample of one area this is to be somewhat expected, when applied to a bigger region with more varied classes it does not translate into useful clusters. In Table 4.7 it is possible to see that some of the clusters' main features coincide, in particular between cluster 0 and cluster 3. Cluster 1 seems to be more related to transportation systems while cluster 2 is the only one that presents a main element directly related to nature (the label "tree").

In Figure 4.7, where k=5, a greater differentiation between clusters is identified. Therefore there are several unique maximum local points (namely for cluster 2 - label "patio" or cluster 3 - label "agriculture"), even if some local maximum and minimum elements still coincide. This variation of peak values can be especially noted in the clusters presented in Table 4.8, where:

- **Cluster 0** does not have any particularly unique feature besides "sky", which

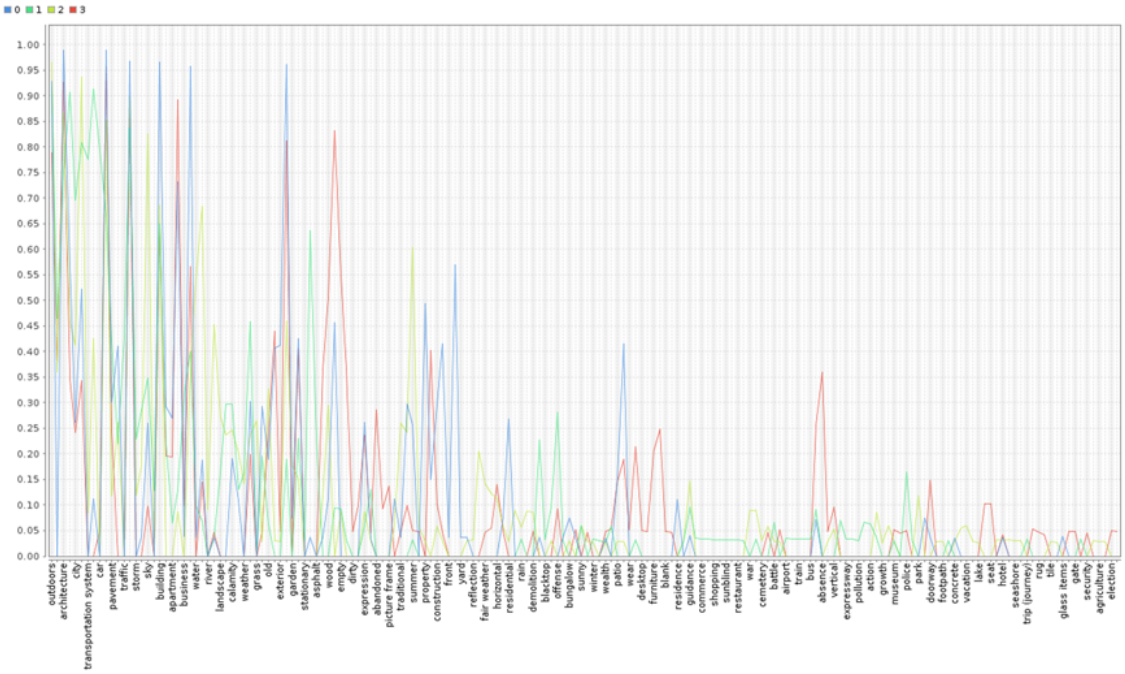


Figure 4.6: Cluster distribution for k=4

Table 4.7: More prevalent cluster characteristics for k=4

Cluster_0	**Cluster_1**	**Cluster_2**	**Cluster_3**
house	outdoors	outdoors	house
architecture	road	travel	architecture
no person	street	no person	window
building	no person	architecture	no person
family	travel	house	door
home	car	sky	family
outdoors	transportation system	building	outdoors
window	architecture	tree	building
	city	summer	
	house		
	building		

could be interpreted as the existence of areas with small buildings where the sky is visible;

- By observing **Cluster 1**, there seems to be a prevalence of transportation related characteristics (“car”, “transportation” and “traffic” are 3 of the eleven main tags);
- Continuing with the analysis, **Cluster 2** is somewhat related to small vegetation in an urban environment (“landscape” and “grass” in conjunction with “window” and “door”. To be noted that, although labels such as “outdoors” also appear with some frequency in this cluster, it is not considered of importance. This is

because when a label appears with a high frequency in a dataset, the weight of its importance in the context is lower. This problem also happens with other indexing techniques such as the Bag-of-words technique. In that case, tf-idf, term frequency–inverse document frequency, is used. Some variation of this term-weighting scheme could be used in future installments to aid in finding the cluster characterization.

- **Cluster 3** seems to be the one with most nature related concepts (“nature”, “tree”, “water”, “flora” and “environment”). This probably means that the images grouped in this cluster will frequently have natural elements, although it is possible to notice some elements related to human construction (for example “car” and “wood”);

- Finally, **Cluster 4** has a lot of common labels, making it hard to identify which are the most discerning ones. The indicator “daylight” is interesting but seems redundant by itself, since all GSV are taken in the daylight. However, it could also be referring to images with high brightness and lack of shadows. Other than “daylight”, the only other concepts of note are “vehicle” and “window”.

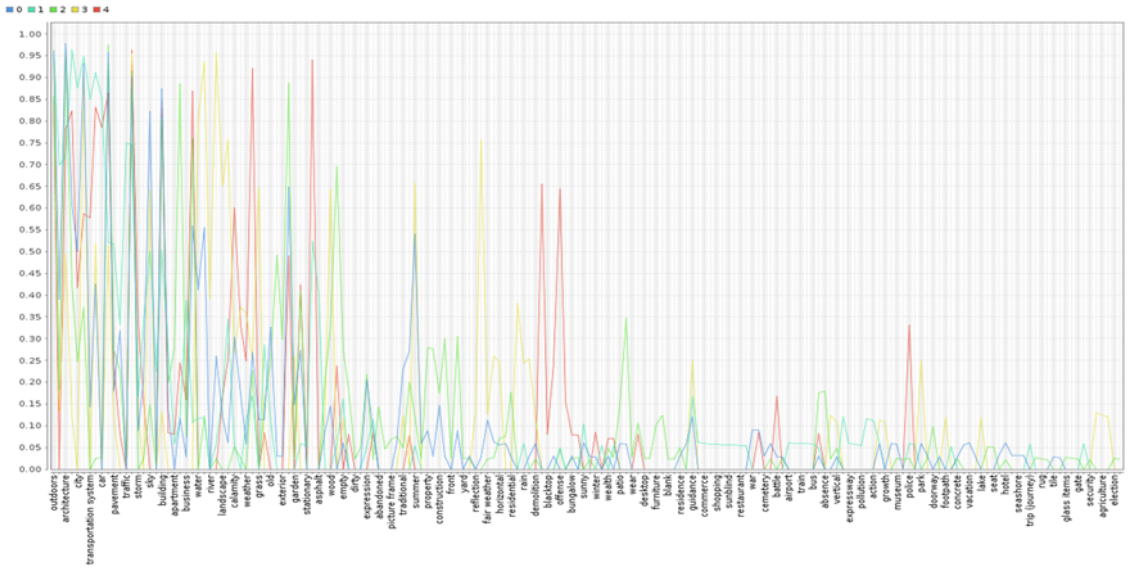


Figure 4.7: Cluster distribution for k=5

When k=6, it is possible to see in Figure 4.8 that there are several different local maximum peaks, and although some overlap, there seems to be a variety that would translate in good clusters. However, as it can be seen in Table 4.9, when comparing results to k=5 there is more repeatability from one cluster to another in different labels that are not the prevailing ones mentioned in Table 4.4. The specific examples are the label “sky” and “road” that appear as the top labels in half of the clusters for k=6. This is to be expected when raising the value of k, but it is not desirable for describing different clusters and also indicates that the clustering model is starting to risk over-fitting the dataset.

Table 4.8: More prevalent cluster characteristics for k=5

Cluster_0	Cluster_1	Cluster_2	Cluster_3	Cluster_4
architecture	street	house	outdoors	no person
outdoors	outdoors	architecture	nature	vehicle
house	travel	no person	no person	daylight
travel	road	family	tree	home
no person	city	window	travel	house
building	car	outdoors	water	building
sky	transportation system	building	flora	outdoors
family	traffic	home	environment	road
street	no person	door	summer	street
	architecture	landscape	car	
	horizontal plane	grass	architecture	
			wood	
			sky	

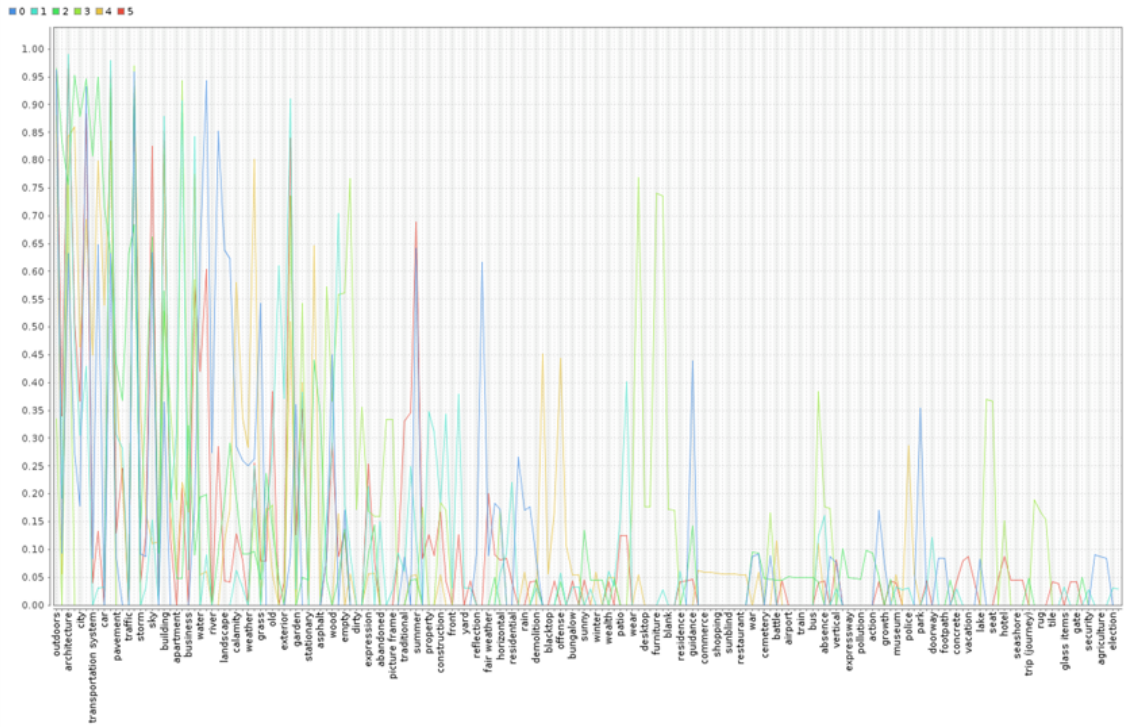


Figure 4.8: Cluster distribution for k=6

Given the previous analysis, the k value of 5 was chosen for the creation of a clustering model sample. One reason is that it offered less overlap between each cluster when compared to a cluster model with a smaller k, such as k = 4. On the other hand, it appeared to offer more differentiated characteristics without starting to overfit, which was observed in cluster model k = 6. Even so, the indicators cannot be directly considered as an index of an area. We need instead to extract from them their main characteristics. Bearing this in mind, and according to the observations previously stated, we can describe each of the five clusters as follows:

Table 4.9: More prevalent cluster characteristics for k=6

Cluster_0	Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5
outdoors	architecture	outdoors	no person	no person	architecture
no person	house	street	window	outdoors	house
tree	outdoors	road	house	street	outdoors
travel	no person	travel	room	architecture	no person
nature	family	city	indoors	house	travel
road	window	horizontal plane	architecture	building	building
water	building	transportation system	furniture	daylight	family
summer	home	architecture	family	road	sky
landscape	door	car	interior design	home	summer
sky	facade	no person	travel	tree	
house		sky		vehicle	
architecture					
environment					
flora					

- **Cluster 0:** presence of human construction and sky;
- **Cluster 1:** presence of transportation methods (e.g. cars, traffic) with a good visibility of the road, since both “road” and “street” appear as main labels;
- **Cluster 2:** presence of buildings with small vegetation (e.g. grass-like or bushes);
- **Cluster 3:** prevalence of nature, flora and sky, and presence of transportation methods;
- **Cluster 4:** presence of transportation methods in an bright area.

For illustration purposes, a random sample of five images for each cluster was chosen as it can be seen in Figure 4.9. Although the characteristics previously presented are represented well, there are some variations within elements. In addition, it is noted that although there were some overlays between the definition of different clusters there are clear disparities between them when observing image examples. A good example of this is observing the differences between images in Cluster 1, classified as transportation methods and a good road visibility, and Cluster 4, transportation methods and bright area. It is also noticed in Cluster 0 that a strong element of that cluster was not noticed in the analysis of the labels, since all images appear to have the presence of Nature in them. A last observation of the different clusters is that the images are overall similar, for example most of them appear to have nature elements, vehicles and buildings. This is to be expected since all images are part of the same area, but it might prove a problem when using this clustering model in other cases where the images are more distinct.

The analysis of the clusters suggests that the clustering method could be improved. Future iterations should perform a deeper study on the influence of some of the parameters of the k-means algorithm. One suggestion is increasing the maximum number of runs. Another suggestion for a future implementation is experimenting with different clustering algorithms even if more computational expensive.

(a) Cluster 0 - Constructions & Sky

(b) Cluster 1 - Transportation methods & Road Visibility

(c) Cluster 2 - Buildings & Small Vegetation

(d) Cluster 3 - Nature & Transportation methods

(e) Cluster 4 - Transportation methods & Open Area

Figure 4.9: Random image samples of clusters when K=5.

4.2.2.1 Cluster Organization comparison with other GSV datasets

After having studied the variations on one dataset example, Sample A, it is interesting to verify what the results are when the same rules are applied to a dataset of the same size but a different location (Cluster C), and the same location but a larger size (Cluster B_1). As such, the purpose of this part is to test the previous methodology on a different environment.

Table 4.10: Sample C from Istanbul region more prevalent cluster characteristics for k=5

Cluster_0	Cluster_1	Cluster_2	Cluster_3	Cluster_4
nature	architecture	no person	architecture	transportation system
tree	house	outdoors	house	outdoors
outdoors	no person	landscape	no person	travel
no person	building	nature	building	road
landscape	outdoors	sky	outdoors	car
flora	travel	travel	street	street
wood	calamity	grass	city	traffic
grass		road	home	no person
environment			travel	vehicle
summer			daylight	city
park			urban	
leaf			family	
travel			window	

The extraction of the features of Sample C using Clarify resulted in 177 different labels. The images where then organized by applying the k-means on Sample C with the same value of k, but with a total number of runs of 1,000 instead of the 60 used before. This difference in the number of runs was due to further experiments which showed a better division of clusters achieved with this value. It is possible to see in Table 4.10 the main labels detected after creating the Sample C cluster model. Like previously discussed, the most characteristic labels, meaning the labels which the number of occurrences in the dataset was bigger than half the dataset size, were determined. By comparing those labels seen in table Table 4.6 with the cluster labels it was easy to define which were the characteristics that better portrayed each cluster like it was done previously for Sample A. The result of that comparison and observation of labels is as follows:

- **Cluster 0:** presence of strong natural elements (out of the 13 labels 9 were related to nature, for example “tree”, “wood” and “leaf”) and an open space (e.g. “landscape”). This cluster does not appear to have buildings;
- **Cluster 1:** Most labels are very common mostly related to the presence of buildings (“architecture” and “houses”) in a possible state of disorder (use of the “calamity” label);
- **Cluster 2:** presence of small vegetation (e.g. “grass”) and some nature near roads in an open space;
- **Cluster 3:** most labels are very similar to cluster 1 but there is also a prevalence of other urban environment’s characteristics without the “calamity” label (e.g.“city”, “urban”) and human constructions (for example “window”);
- **Cluster 4:** presence of transportation methods like “car” and “traffic” in an urban environment (“city”).

Again, for illustration purposes, a small sample of 5 images was randomly chosen to illustrate the clusters, seen in Figure 4.10. Here it is possible to see that although some clusters, and the images present in them, are very distinct, there are characteristics of others that overlap to a certain degree. For example, green areas are supposed to appear mainly on cluster 0 and 2, yet on the first image of cluster 1 there is a great area of small vegetation. In cluster 3, there are also green areas, although one could argue they are present in the characteristic urban environment of that cluster. An interesting observation in this random sample is comparing the last image of Cluster 3 and 4. Both these images include a vehicle in roughly the same position, yet the biggest clear difference between the two images is the existence or lack thereof of construction or buildings. This distinction verifies the previously expressed possible outcome from the main labels identified in cluster 3.

By repeating, in Sample C, the same methodology used to organize Sample A, it was possible to identify two distinct clustering profiles between the two regions. As such, the next study to be done was on the results of **applying the Sample A clustering model in the Sample C dataset**. To do this, the clustering model was trained using Sample A and applied to Sample C to achieve results. An example of

(a) Cluster 0 - Strong Nature & Open Space

(b) Cluster 1 - Buildings & Disorder

(c) Cluster 2 - Road & Small Vegetation

(d) Cluster 3 - Urban Environment & Constructions

(e) Cluster 4 - Transportation methods & Urban Environment

Figure 4.10: Random image samples of clusters when K=5 for Istanbul region.

(a) Cluster 0 - Constructions & Sky

(b) Cluster 1 - Transportation methods & Road Visibility

(c) Cluster 2 - Buildings & Small Vegetation

(d) Cluster 3 - Nature & Transportation methods

(e) Cluster 4 - Transportation methods & Open Area

Figure 4.11: Example images of applying the clustering model resulting from Sample A to Sample C dataset.

the image classification according to this method can be seen in Figure 4.11. Further studies on this result can be seen in the result visualization section.

4.2.3 Visualization Results

Visualization is made through the website and through the analysis of data of the administrative division area. Both of these approaches are considered in this module because both work from the coordinates regardless of the indicators received.

The website extracts processed information and its coordinates according to the previously created database. The website dashboard consists of the display of a map where it is possible to select one of four options and obtain the places where those characteristics are more prominent. An example of the implementation can be seen in Figure 4.12. To be noted that this technique allows for the inspection of an area in a way more suitable for users without experience with data analysis.

The other technique for the visualization of results is achieved by getting the coordinates of the interest points and extracting from them the address, the district,

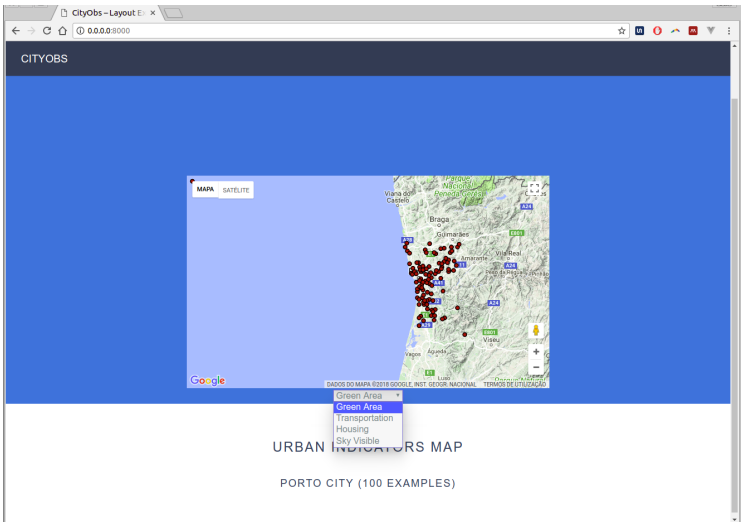


Figure 4.12: Dashboard Visualization Example.

municipality, parish and postal code. An issue verified in this extraction was that the Google Street View's geocoding API does not always send completely accurate information, meaning that in some cases it lacks information on details such as municipality and parish. This complicates an accurate analysis on the characteristics of those administrative divisions. A possible correction would be getting coordinates of the delimitations from official sources.

To do this, an extraction of the administrative division was applied to sample A. In order to facilitate comprehension, the data was divided according to the previously defined clusters. Using the existent data, it was decided to compare the profile of Porto and Aveiro according to the clustering previously created. However, since this sample size is small, this study works mostly as a proof of concept. The exact sample for each district was a total of 68 images of Porto and 25 images of Aveiro. The conclusion of the analysis based on this sample is that Porto has bigger areas of construction and visible sky areas, while Aveiro has a greater prevalence of nature and transportation methods like it can be seen on Figure 4.13.

In regards to sample C, it is also interesting to visualize the results of the cluster organization. After having studied the clusters and seeing the particular examples of each cluster in Section 4.2.2.1, it is interesting to see their distribution in the region. To be noted that, just like the previous study of the Porto Region included samples from other surrounding areas, this general distribution of the area of Istanbul will also include some neighboring areas.

It can be seen in Figure 4.14 that there is a considerable area with a clear nature presence and almost no buildings. It can also be observed that a quarter of the region is road and small vegetation. Although 28%, more than one quarter, is classified as urban environment and constructions, there still seems to be a high prevalence of green areas.

There is a need to verify this assumption on the indicators of Region C raised by its clustering organization profile. For this reason, a direct comparison was performed

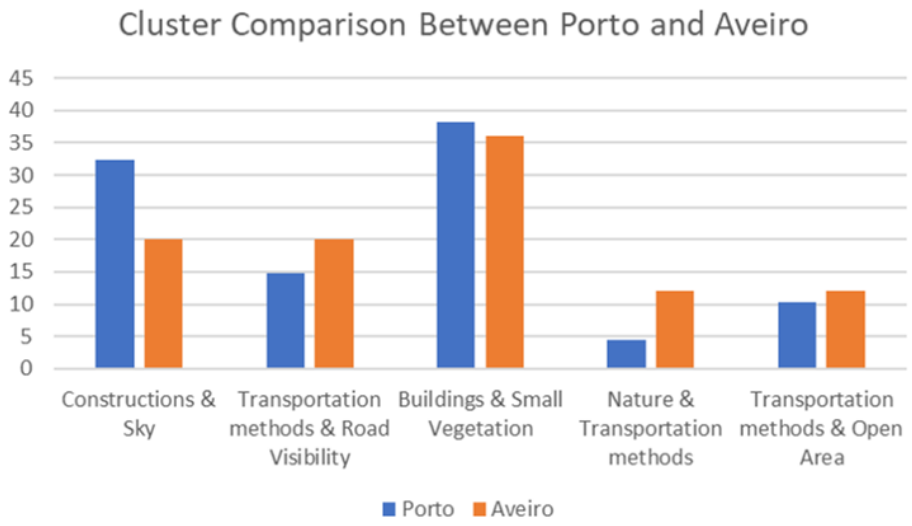


Figure 4.13: Comparison between two different districts from Sample A, Aveiro and Porto using Sample A Cluster Model.

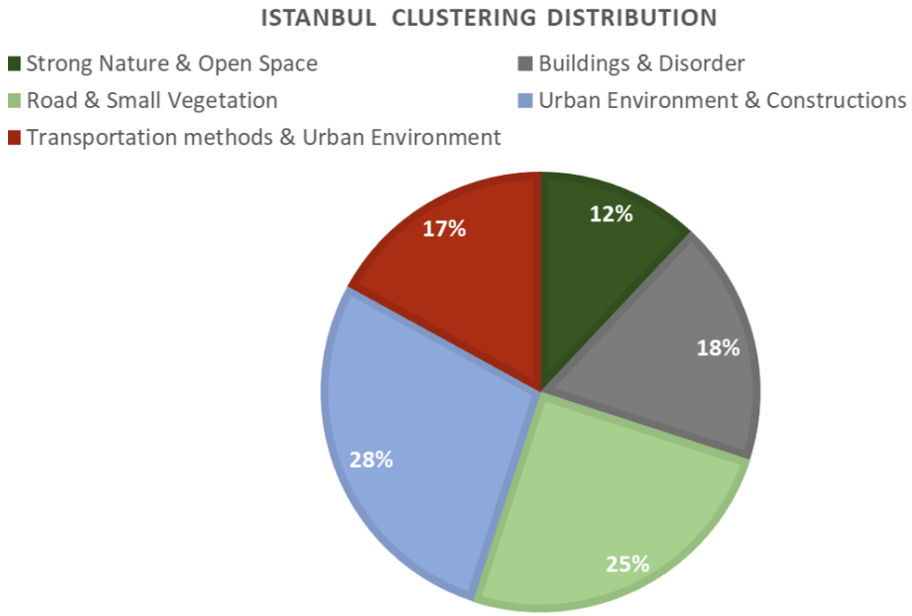


Figure 4.14: Data Clustering of the Istanbul Region

between Region A and C, both using the Sample A clustering model. The visualization of these results can be seen in Figure 4.15.

As predicted, Sample C has a strong prevalence of Nature, at least in relation to Sample A. However, it is noticeable that region A also has a similar peak of percentage of elements, in comparison to Sample C, in the cluster Buildings and Small Vegetation. Sample A in this evaluation has a bigger percentage of the Constructions and Sky category. However, Sample C seems to have a higher prevalence of Transportation methods by looking as a whole to the distribution of the different clusters. By adding

these values one reaches the percentage of 71% of the elements.

By observing these calculations it is possible to draw the hypothesis that Sample C has a high prevalence of transportation methods. Which is wrong, when observing the previous profile, seen in Figure 4.10. Here, unlike when using the Sample A clustering model organization, only one of the categories of Sample C is characterized by the prevalence of transportation methods, representing about 17% of the whole sample.

This observation reiterates the fact that the clustering model from Sample A needs improvements either by a better selection of the parameters of k-means or by experimenting with other clustering methods. To be noted that the seemingly better organized clustering model from sample C had a larger maximum number of runs.

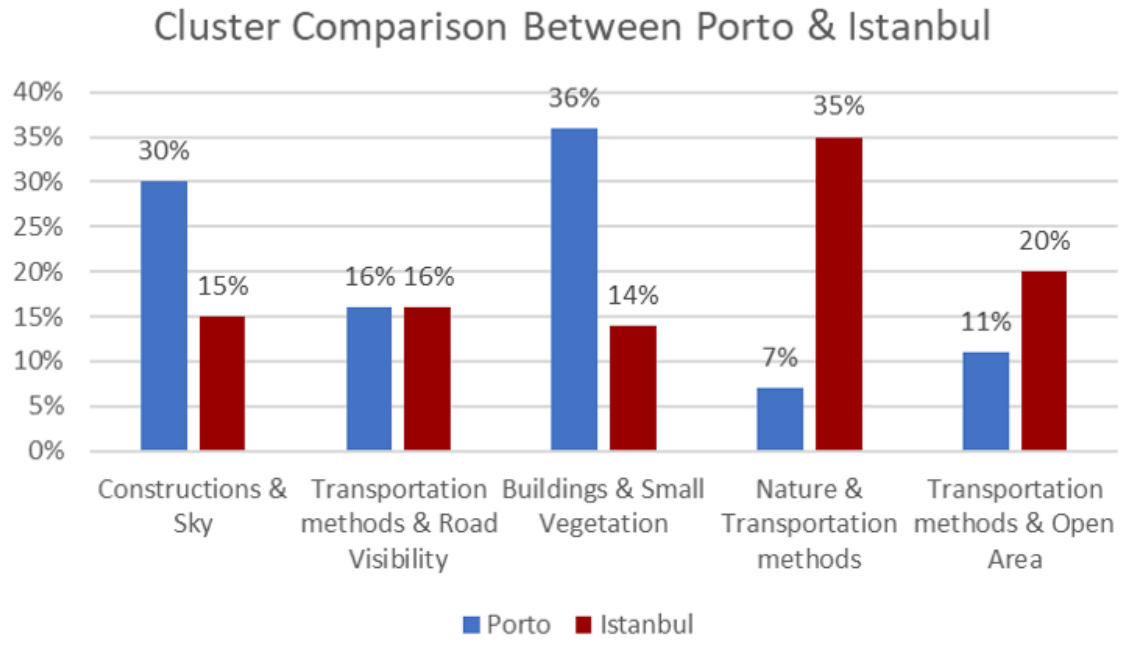


Figure 4.15: Comparison between Sample A (Porto) & Sample C (Istanbul) regions using Sample A Cluster Model(%).

Alternative Datasets

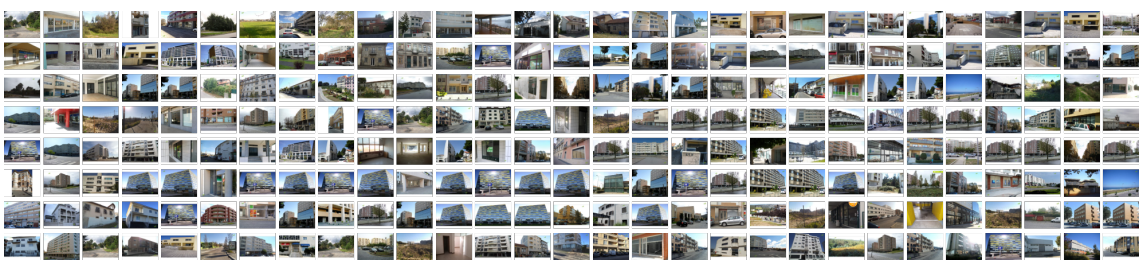


Figure 4.16: Dataset extracted from real estate website for Porto region.

Another dataset was obtained by using the crawler presented in the implementation section. It needs improving for optimal usage, but it shows potential for future iterations. The execution of this crawler resulted in a dataset containing data on 232 images of size 280x210px which can be seen in Figure 4.16 in the region of Porto. This dataset provided new insights into an alternative means of obtaining information about the city, for example, by relating the visual appearance of a facade of a building to its price.

A short examination was made regarding this dataset. We saw that the available data on the prices varied between 4,000 to 5,879,500€, being that the lowest value corresponded to garages and the highest to land for sale in the central part of the city. Another conclusion from the results was the variation of price which can be seen can be seen in Figure 4.17. To be noted that, out of the 232 samples, 4 were classified as "after contract" and 3 had a variation of price proposed, possibly because they were labeled as "offices" or "shops", and were not considered for the analysis.

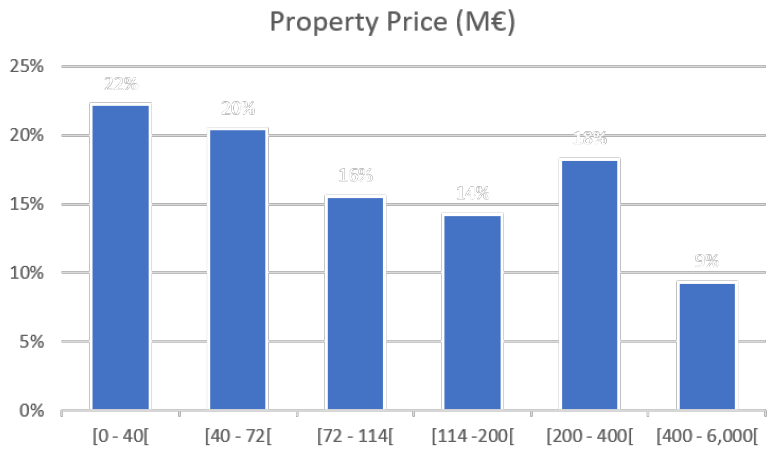


Figure 4.17: Variations of Property Prices for sale in Porto region.

Although this alternative dataset was not extensively used in the main implementation of this project, it will be saved and used for future developments. In particular, one very interesting study could be made from the "year of the construction" label and the image of the facade of the house. This might result in a classification model capable of predicting the original year of construction of new datasets. This conjecture is made in the context of the planning rules, hight costs and bureaucracy usually constraining structural changes to the outer appearance of a house. Meaning that the facade is a characteristic subject to little changes over time and as such, ideal to be evaluated using GSV imagery.

However, the dataset characteristics bear possible challenges before being used to train a model, such as:

- Watermark - the images contain a small watermark that may cause complications when extracting features. It might be possible to remove them automatically, but further research on that would have to be made;

- Point of view from where the image is taken - the images are not from GSV and as such there might be differences between the point of view the photographs were taken from, and thus not containing an adequate view of the facade, which might proceed to create incorrect results;
- Frame - some of the images have a small, white, horizontal or vertical frame, so it is advisable to cut those areas before proceeding to the feature extraction.

4.3 Summary

This chapter was focused on the implementation of the modules and the description, exposition and discussion of the main results of this dissertation. The implementation of the modules did not cover all the proposed methodologies in Chapter 3, but it did build the desired output to ascertain conclusions on the visual urban indicators extracted. It was possible to conclude that 9 in 178 labels returned from the 100 images sample were too common to be considered useful for comparison between areas. The results of applying clustering, using unsupervised Machine Learning algorithms, to profile the different indicators according to groups suggest that it is possible to obtain human understandable patterns in an automated fashion. It was even possible to remark on some characteristics in different districts of the Region of Porto.

The comparison between different Samples showed that the clustering model from Sample A still needs improvements which can be achieved by testing more parameters with the the k-means methods. The proof of that was Sample C clustering model, while changing the parameter of the maximum number of runs from 60 to 1000, it was capable of having more clearly differentiated clusters. Another alternative can be testing other unsupervised clustering methods more computationally expensive.

Also, the usage of the crawler to extract a dataset from official classifications on real estate websites seems to be very promising, as it is expected to allow for the classification of the houses of a section of a city according to their characteristics. Finally, there is room for future different implementations of the modules, in particular training the model for image classification, by using the alternative dataset extracted from real estate websites.

Chapter 5

Conclusions and Future Work

5.1 Main Considerations

To evaluate the characteristics of an urban environment, methods to ascertain on its sustainable and effective growth are crucial. In that context, the inference of urban indicators using alternative, more automated, standardized and low-resource-dependent methods provides a clear advantage. This is especially true if we consider that the extraction of visual indicators suffers from several limitations; for example, secondary analysis of archival data sources that do not have a standard data organization, survey-based measures may be influenced by subjectivity and source bias, and the use of empirical audit instruments can be highly resource-intensive and costly. Thus, this dissertation studied the necessary foundations and implemented a way of extracting urban indicators — in this particular case, visual urban indicators — recurring to Google Street View and Computer Vision techniques. The goal of this work was to extract information, organize the results and parametrize data into intuitive and meaningful structures.

To better understand the concepts behind this type of extraction, a literature review was conducted, using a systematic literature review process as the main guideline. However, due to issues possibly related to choosing not-so-good keywords, the data extraction and qualification was not completely satisfactory, particularly when obtaining precise definitions of concepts was the main objective of the survey carried out. Nonetheless, a number of related references were identified and discussed in the work, therefore providing use with the necessary background to devise the proposed methodology.

Urban indicators are the metrics capable of providing information on the urban performance in various dimensions. Considering the multitude of different urban contexts in one region, as well as all over the world, there are naturally great varieties of possible indicators. Most of those are related to three main dimensions of life in a urban area, namely the Environment, the Society, and the Economy. Yet, the relevance of each indicator is hard to infer upon. The proposed approach to this problem is to choose urban indicators that are able to provide information on a specific aspect of the environment, such as its sustainability. In the particular case of visual urban indicators, it has not yet been found a global classification that would differentiate them from others. Therefore, in the proposed approach to this dissertation scope

we use existing structures to organize and devise a taxonomy of viable automated processes.

Computer Vision still faces several challenges such as recognition of desirable features, understandable data output, segmentation, and reasonable and balanced use of computational resources. Nonetheless, the evolution of the Machine Pattern Recognition field and, as such, the evolution of Computer Vision in general, have made it possible for these techniques to be considered capable of providing advantage grounds to the development of automatic exploration of GSV imagery in an urban context. Taking that into account, this dissertation proposes the following three main hypotheses:

- It is possible to organize visual urban indicators according to a given taxonomy;
- There are already appropriate and available data easily accessible on GSV tools;
- Deep learning algorithms are promising classification approaches that may support a successful process for extracting at least high level features from the GSV imagery.

A proposed methodology was planned accordingly, taking into account the possible benefits of a tool structured into the main modules identified. The objective of such an structure is to solve the complexity of inferring urban indicators from GSV. The phases identified were the extraction of the necessary dataset, the classification of such dataset, and finally the visualization of the urban indicators inferred from the data. The dataset uses images extracted from Google Street View and, in this case study, accounts for the selection of the area to be analyzed in terms of boundaries, diversity of that area's characteristics and availability of other information sources so as to validate the data extracted.

The classification of the dataset identifies characteristic of the ideal dataset for training a classification model. It also envisions other classification methods that have lesser need for computational resources by using an external image recognition application programming interface. The use of unsupervised clustering techniques was discussed for automating the creation of a taxonomy for the urban indicators. This clustering would allow a standard comparison between different regions.

The visualization module was intended to provide a general straightforward interpretation of urban indicators, capable of displaying the classification results through different visual metaphors. To achieve this, the proposed optimal methodology would consist of a visualization dashboard and the use of graphical diagrams, plots, and charts, providing users with a visual interpretation of indicators. The graphic diagrams would be used to evaluate regions at the level of administrative divisions, allowing for a means of comparing two geographically distant regions according to various dimensions.

From the implementation of these modules, it was possible to build a positive output to ascertain conclusions on the visual urban indicators extracted. The frequency of some classification labels extracted by Clarifai was deemed too high to be effective in differentiating clusters. A prospection exploration of different tool for image recognition, possibly considering combinations of methods, is certainly crucial adjust their parameters and response values. A proposed future improvement uses Bag of words'

tf-idf technique to automatically weight and organize the labels in a more effective way, for instance.

The results of clustering using unsupervised Machine Learning algorithms to profile the different indicators according to groups suggested that it was also possible to obtain human understandable patterns automatically. However, some improvements are also possible when using k-means in the first iteration of the implementation of the modules. Sample A displayed some flaws in its clustering, since some characteristics were very common in different clusters, as compared to Cluster C, which had a higher number of maximum runs of k-means in the clustering algorithm. In spite of that, it was still possible to infer some characteristics in different regions based on the different clustering model profiles. A small study on Cluster B_1 was also made, whose region is the same as Sample A but with a dataset 10 times bigger.

Finally, during the development phase, new ideas for for alternative implementations of the modules came up. In particular, it would be of interest to train the model for image classification, by using the alternative dataset extracted from real estate websites. The dataset with 233 images and prices of properties near the region of Sample A was extracted for future studies. The images however have some limitations, such as the existence of watermarks (in this case a small green mark is present in all images). For this reason, future utilizations of this dataset need to account for further considerations and appropriate techniques to address the aforementioned limitations.

5.2 Further Developments

Bearing in mind this is an initial exploratory approach, there are several aspects requiring further study, investigation, and development, and that may represent interesting topics for future research projects. We discuss some of such ideas as follows.

Improvement of clustering algorithm — Further testing of the algorithms should be done in order to find even better parameters namely for the k-means technique; both the number of k when applied to a bigger dataset and the optimal maximum number of runs need to be tuned up. Another improvement to the clustering description for human visualization would imply to formally apply the Bag of words' tf-idf metrics to the list of labels.

Elaboration of a classification model from the existing extracted dataset — There are currently two types of datasets that are extracted in this methodological approach, namely the samples extracted by Google Street View that were used for extracting the urban indicators, and the image samples collected from the real estate website. Both have the potential to be used as a dataset for training a classification model. Further work on these datasets however are necessary in order to improve classification results, as follows.

- Google street view dataset – manual labeling of the sample images by specialists in the field of urban development;

- Real estate image dataset - conduct tests to the images in order to see how appropriate they are for the classification model (i.e. we need to evaluate the

effect of a number of limitations of this dataset on the classification results, such as the consequence of the green watermark present in all images).

Combine external recognition tools — Although using Clarifai as an image recognition tool allowed us to achieve the extraction of urban indicators, each external image recognition tool yields a different set of results. Hence it would be interesting to combine them and see what other data organizations would possibly emerge (e.g. different clusters or relationships perhaps).

Extend the dataset from real estate websites — Extending the present dataset including images crawled from a broader group of real estate websites in different regions will widen the amount and quality of information. Certainly this will imply to deal with images presenting different qualities, which demands for more robust techniques. Nonetheless, using multiple and diverse sources enables us to do different studies on urban characteristics related to housing and construction as well.

Study of more extensive databases — Performing the extraction, classification and visualization of more points inside of a region certainly contributes to the generation of better profiles of an area. Thus it would be useful to implement a means to identify the minimum number of points per area that would allow for an accurate profile of a region. Such a study can greatly contribute to the optimum use of computational resources, having important impact on performance.

Besides the necessary improvements to the methodological approach herein proposed, other future work efforts can originate from this project, adding new services on top of the devised reference architecture and contributing to other applications as well.

The marketing on real estate areas — Further studies on the urban indicators extracted by computer vision will allow for a classification of areas according to the profile of the buildings there present. This leverages the marketing of those areas with more appropriate mechanisms to promote different products to potential costumers.

Recommendation systems — By creating an open source dashboard tool to visualize geographic urban environments it is possible to help users to find the most suitable area to live or to visit. This certainly have applications to the real state marketplace as well as to tourism. Thus the ability to identify the characteristics that make an area appealing and finding places that bear the same characteristics is an important ingredient in recommendation systems. Such tools present interesting potential to find new places that users would never have considered otherwise.

Support of territorial management, land use and land cover, zoning, as well as decision-making in the public sector — Public decision makers in charge of managing the territorial aspects of a city or a region demand for appropriate decision support tools to characterize different aspects of the territory, regarding land use and land cover, zoning systems, property records, and so forth. The extraction of urban indicators by using this tool might represent an important asset to help decision making in the public sector. For example, to quickly assess the density of buildings present in a region so as to decide upon the placement of waste treatment plants, and to optimize the placement of garbage collection containers over the city are certainly activities that can greatly benefit from the proposed system.

Identification of fire risk areas — With the current dataset it is not possible to extract information beyond the existence of buildings and strong nature elements in the vicinities of streets and roads. However by expanding the dataset to forest areas and additionally classifying it through a vegetation classification model it is possible to implement notifications of dangerous areas close to buildings. This approach takes advantages of the Google Street View's perspective from which it is possible to see the kind of vegetation and evaluate its conditions; for instance, areas where there are trimmed trees reduce the chance of fire climbing the threes and grow uncontrolled. However, distinguishing differences in green areas poses lot more challenges; robust techniques are necessary to cope with the complexity of identifying useful features in images in which shapes of objects are not easily recognizable. Albeit not in the scope of this dissertation, nevertheless representing an area to which the proposed methodology is certainly applicable, the use of satellite or aerial survey imagery also represent a promising approach. A quadcopter platform is currently under development to capture images of streets and motorway for congestion analysis [VOP+14, VKP+14]. Such images could be later used as an additional source in our approach as well.

Standardization of Urban Indicators on a global scale — It would be interesting to apply the current modules to a worldwide area in order to find out the kind of clusters that would form up. To be noted that according to the proposed approach in this dissertation the areas to cover would be limited. Such a limitation is inherent to the constraints imposed by the current technology of Google Street View (only tracks, streets, and roads on which we can navigate are included), and to local government regulations of each country. Cultural aspects also apply in such a worldwide perspective. Indeed, some indicators can have different interpretations in different countries. Also, the required effort to label sufficient examples would be tremendous, requiring crowdsourcing or other similar methods to enlist the participation of a large number of people.

Finally, this work presents direct potential contributions to the MAS-Ter Lab framework [ROB07] under development at LIACC, University of Porto. It naturally extends the dashboard initially devised to monitor mobility metrics [ZRC14] in combination with a wider perspective of urban management [RRC17] to enhance the analytical capabilities of MAS-Ter Lab. The possibility of combining the automatic inference of visual indicators with opinion mining and sentiment analysis over social networks, such as studies initiated elsewhere [CSR10, RSR15, USRS16, PPS+17], represents a huge step ahead towards the consolidation of an integrated platform to monitor, manage, and evolve smarter cities and smarter societies.

References

[AAMC18] Mahmood Abdulkareem, Sura Al-Maiyah, and Malcolm Cook. Remodelling façade design for improving daylighting and the thermal environment in Abuja's low-income housing. *Renewable and Sustainable Energy Reviews*, 82:2820–2833, 2 2018.

[AT13] Alexander Andreopoulos and John K. Tsotsos. 50 Years of object recognition: Directions forward. *Computer Vision and Image Understanding*, 117(8):827–891, 8 2013.

[BGdVL] G. Bebis, M. Georgiopoulos, and N. da Vitoria Lobo. Learning geometric hashing functions for model-based object recognition. In *Proceedings of IEEE International Conference on Computer Vision*, pages 543–548. IEEE Comput. Soc. Press.

[BHK+14] Marc G Berman, Michael C Hout, Omid Kardan, MaryCarol R Hunter, Grigori Yourganov, John M Henderson, Taylor Hanayik, Hossein Karimi, and John Jonides. The perception of naturalness correlates with low-level visual features of environmental scenes. *PloS one*, 9(12):e114572, 2014.

[BOW+10] Hannah M. Badland, Simon Opit, Karen Witten, Robin A. Kearns, and Suzanne Mavoa. Can Virtual Streetscape Audits Reliably Replace Physical Streetscape Audits? *Journal of Urban Health*, 87(6):1007–1016, 12 2010.

[BPT94] S.J. Broadhurst, T.P. Pridmore, and N. Taylor. Sensing for feature identification in sewers. In *Automation and Robotics in Construction Xi*, pages 675–682. Elsevier, 1994.

[Bra91] Leon Braat. The predictive meaning of sustainability indicators. In *In Search of Indicators of Sustainable Development*, pages 57–70. Springer Netherlands, Dordrecht, 1991.

[CDF+04] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.

[CH67] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1 1967.

[CLY15] Yi-Ting Chen, Xiaokai Liu, and Ming-Hsuan Yang. Multi-instance object segmentation with occlusion handling. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3470–3478. IEEE, 6 2015.

[COR+16] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4 2016.

[CSCC16] Iwona Cieslak, Karol Szuniewicz, Szymon Czyza, and Cieslak I Szuniewicz K Czyza S. Analysis of the Variation of the Areas Under Urbanization Pressure Using Entropy Index. *Procedia Engineering*, 161:2001–2005, 2016.

[CSR10] Sara Carvalho, Luís Sarmento, and Rosaldo J. F. Rossetti. Real-time sensing of traffic information in twitter messages. In *4th Workshop on Artificial Transportation Systems and Simulation (ATSS), 2010 13th International IEEE Conference on Intelligent Transportation Systems - (ITSC 2010), Funchal, Portugal, 19-22 Sept. 2010*, pages 1–4, 2010.

[CV95] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 9 1995.

[Dav12] Roy Davies. *Computer and Machine Vision, 4th Edition Theory, Algorithms, Practicalities Opsylum*. 2012.

[DF11] Thomas Deselaers and Vittorio Ferrari. Visual and semantic similarity in ImageNet. In *CVPR 2011*, pages 1777–1784. IEEE, 6 2011.

[DGL96] Luc Devroye, László Györfi, and Gábor Lugosi. *A Probabilistic Theory of Pattern Recognition*, volume 31 of *Stochastic Modelling and Applied Probability*. Springer New York, New York, NY, 1996.

[DH72] Richard O. Duda and Peter E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1 1972.

[Eco12] Economist Intelligence Unit. *The Green City Index*. Siemens AG, Munich, 2012.

[EG14] Ian M. S. Eddy and Sarah E. Gergel. Why landscape ecologists should contribute to life cycle sustainability approaches. *Landscape Ecology*, 30(2):215–228, 2 2014.

[Fak16] Ioanna Fakiri. The Interpretation of Landscape - Strategies Towards Smarter Cities. *Proceedings of the 5th International Conference on Smart Cities and Green ICT Systems*, pages 111–116, 6 2016.

[FCR+16] T. Feuillet, H. Charreire, C. Roda, M. Ben Rebah, J. D. Mackenbach, S. Compernolle, K. Glonti, H. B?rdos, H. Rutter, I. De Bourdeaudhuij,

M. McKee, J. Brug, J. Lakerveld, and J.-M. Oppert. Neighbourhood typology based on virtual audit of environmental obesogenic characteristics. *Obesity Reviews*, 17(S1):19–30, 1 2016.

[Gal97] G.C. Gallopin. Indicators and their use: information for decision-making. *In: Moldan, B., Billharz, S. (Eds.), Sustainability Indicators.*, 58:13–27, 1997.

[Gir15] Ross Girshick. Fast R-CNN. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448. IEEE, 12 2015.

[GJM+11] Francisco Gómez, José Jabaloyes, Luis Montero, Vicente De Vicente, Manuel Valcuende, Francisco G?mez, José Jabaloyes, Luis Montero, Vicente De Vicente, Manuel Valcuende, Francisco Gómez, José Jabaloyes, Luis Montero, Vicente De Vicente, and Manuel Valcuende. Green Areas, the Most Significant Indicator of the Sustainability of Cities: Research on Their Utility for Urban Planning. *Journal of Urban Planning and Development*, 137(3):311–328, 9 2011.

[GJS+17] Robert Geirhos, David H J Janssen, Heiko H Schütt, Jonas Rauber, Matthias Bethge, and Felix A Wichmann. Comparing deep neural networks against humans: object recognition when the signal gets weaker. 2017.

[GL11] Kristen Lorraine Grauman and Bastian. Leibe. *Visual object recognition*. Morgan & Claypool Publishers, 2011.

[Goo17a] Google. Where We've Been And Where We're Headed Next – Google Street View, 2017.

[Goo17b] Google Maps. Praça do Marquês de Pombal - Google Maps, 2017.

[HAGM14] Bharath Hariharan, Pablo Arbeláez, Ross Girshick, and Jitendra Malik. Simultaneous Detection and Segmentation. pages 297–312. Springer, Cham, 2014.

[Har15] Bharath Hariharan. *Beyond Bounding Boxes: Precise Localization of Objects in Images | EECS at UC Berkeley*. PhD thesis, EECS Department, University of California, Berkeley, 2015.

[HEH+08] Derek Hoiem, Alexei A Efros, Martial Hebert, D Hoiem, A A Efros, · M Hebert, and M Hebert. Putting Objects in Perspective. *Int J Comput Vis*, 80:3–15, 2008.

[HEHE07] James Hays, Alexei A. Efros, James Hays, and Alexei A. Efros. Scene completion using millions of photographs. *ACM Transactions on Graphics*, 26(3):4, 7 2007.

[HEL+14] A. Hsu, J. Emerson, M. Levy, A. de Sherbinin, L. Johnson, O. Malik, J. Schwartz, and M. Jaiteh. The 2014 Environmental Performance Index. Technical report, Yale Center for Environmental Law and Policy, New Haven, CT, 2014.

[HLF13] Kotaro Hara, Vicki Le, and Jon Froehlich. Combining crowdsourcing and google street view to identify street-level accessibility problems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, page 631, New York, New York, USA, 2013. ACM Press.

[HRA+17] Roger Hyam, J Roe, P Aspinall, R Mitchell, A Clow, and D Miller. Automated Image Sampling and Classification Can Be Used to Explore Perceived Naturalness of Urban Spaces. *PLOS ONE*, 12(1):e0169357, 1 2017.

[HU90] Daniel P. Huttenlocher and Shimon Ullman. Recognizing solid objects by alignment with an image. *International Journal of Computer Vision*, 5(2):195–212, 11 1990.

[HWY15] Lu Huang, Jianguo Wu, and Lijiao Yan. Defining and measuring urban sustainability: a review of indicators. *Landscape Ecology*, 30(7):1175–1193, 8 2015.

[JMF+99] A K Jain, M N Murty, P J Flynn, Azriel Rosenfeld, K Bowyer, N Ahuja, and A Jain. Data Clustering: A Review. *ACM Computing Surveys*, 31(3), 1999.

[KA59] L. A. Kamentsky and L. A. Pattern and character recognition systems. In *Papers presented at the the March 3-5, 1959, western joint computer conference on XX - IRE-AIEE-ACM '59 (Western)*, pages 304–309, New York, New York, USA, 1959. ACM Press.

[KBS15] Virendra Kumar, Kamlesh Bhalvai, and Anugya Shukla. Quantification of land transformation using multi temporal satellite data and GIS techniques. In *2015 National Conference on Recent Advances in Electronics & Computer Engineering (RAECE)*, pages 106–111. IEEE, 2 2015.

[KFF17] Andrej Karpathy and Li Fei-Fei. Deep Visual-Semantic Alignments for Generating Image Descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):664–676, 4 2017.

[Kit04] Barbara Kitchenham. Procedures for Performing Systematic Reviews. Technical report, Department of Computer Science, Keele University, UK, Keele, UK, 2004.

[KP14] Anders Kofod-Petersen. How to do a Structured Literature Review in computer science. 2014.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks, 2012.

[KSvM+17] Monika Kuffer, Richard Sliuzas, Martin van Maarseveen, Karin Pfeffer, and Isa Baud. City nighttime light variations using ISS images. In *2017 Joint Urban Remote Sensing Event (JURSE)*, pages 1–4. IEEE, 3 2017.

[KT02] A. M. Kabayama and L. G. Trabasso. Performance evaluation of 3D computer vision techniques. *Journal of the Brazilian Society of Mechanical Sciences*, 24(3):234–238, 7 2002.

[LKR+16] G. Lira, Z. Kokkinogenis, R. J. F. Rossetti, D. C. Moura, and T. Rúbio. A computer-vision approach to traffic analysis over intersections. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 47–53, Nov 2016.

[LRB09] P. F. Q. Loureiro, R. J. F. Rossetti, and R. A. M. Braga. Video processing techniques for traffic information acquisition using uncontrolled video streams. In *2009 12th International IEEE Conference on Intelligent Transportation Systems*, pages 1–7, Oct 2009.

[LRdSM16] Patricio Loncomilla, Javier Ruiz-del Solar, and Luz Martínez. Object recognition using local invariant features for robotic applications: A survey. *Pattern Recognition*, 60:499–514, 12 2016.

[LW04] Dengsheng Lu and Qihao Weng. Spectral Mixture Analysis of the Urban Landscape in Indianapolis with Landsat ETM+ Imagery. *Photogrammetric Engineering & Remote Sensing*, 70(9):1053–1062, 9 2004.

[MC12] Koichiro Mori and Aris Christodoulou. Review of sustainability indices and indicators: Towards a new City Sustainability Index (CSI). *Environmental Impact Assessment Review*, 32(1):94–106, 1 2012.

[MC15] Priyanka Mukhopadhyay and Bidyut B. Chaudhuri. A survey of Hough Transform. *Pattern Recognition*, 48(3):993–1010, 3 2015.

[MCN+] Paula Mian, Tayana Conte, Ana Natali, Jorge Biolchini, and Guilherme Travassos. A Systematic Review Process for Software Engineering. *ESELAW'05: 2nd Experimental Software Engineering Latin American Workshop*.

[MH05] M Katherine McCaston and HLS Advisor. Tips for Collecting, Reviewing, and Analyzing Secondary Data, 2005.

[MVdV+09] J Maas, R A Verheij, S de Vries, P Spreeuwenberg, F G Schellevis, and P P Groenewegen. Morbidity is related to a green living environment. *Journal of epidemiology and community health*, 63(12):967–73, 12 2009.

[NPRH14] Nikhil Naik, Jade Philipoom, Ramesh Raskar, and Cesar Hidalgo. Streetscore – Predicting the Perceived Safety of One Million Streetscapes. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 793–799. IEEE, 6 2014.

[OFT+09] Åsa Ode, Gary Fry, Mari S. Tveit, Pernette Messager, and David Miller. Indicators of perceived naturalness as drivers of landscape preference. *Journal of Environmental Management*, 90(1):375–383, 1 2009.

[Ole06] Nancy Olewiler. Environmental sustainability for urban areas: The role of natural capital indicators. *Cities*, 23(3):184–195, 6 2006.

[OT01] Aude Oliva and Antonio Torralba. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope *. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[Pap66] Seymour Papert. *The summer vision project*. Massachusetts Institute of Technology Project MAC, Cambridge Mass., 1966.

[PC16] Margherita Pongiglione and Chiara Calderini. Sustainable Structural Design: Comprehensive Literature Review. *Journal of Structural Engineering*, 142(12):04016139, 12 2016.

[PHD] V. Philomin, D. Harwood, and L.S. Davis. Appearance-based automatic target recognition in overhead LADAR range imagery. In *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*, volume 2, pages 1320–1324. IEEE Comput. Soc.

[PPS+17] J. Pereira, A. Pasquali, P. Saleiro, R. Rossetti, and N. Cacho. Characterizing geo-located tweets in brazilian megacities. In *2017 International Smart Cities Conference (ISC2)*, pages 1–6, Sept 2017.

[PPSR17] João Pereira, Arian Pasquali, Pedro Saleiro, and Rosaldo Rossetti. Transportation in social media: An automatic classifier for travel-related tweets. In Eugénio Oliveira, João Gama, Zita Vale, and Henrique Lopes Cardoso, editors, *Progress in Artificial Intelligence. EPIA 2017. Lecture Notes in Computer Science*, volume 10423 of *LNCS*, pages 355–366, Cham, 2017. Springer International Publishing.

[PVG+01] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. *Journal of machine learning research : JMLR.*, volume 12. MIT Press, 2001.

[PZZ13] Bo Peng, Lei Zhang, and David Zhang. A survey of graph theoretical approaches to image segmentation. *Pattern Recognition*, 46(3):1020–1038, 3 2013.

[QD14] Daniele Quercia and Daniele. The pursuit of urban happiness. In *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*, pages 611–612, New York, New York, USA, 2014. ACM Press.

[QMS+16] James W. Quinn, Stephen J. Mooney, Daniel M. Sheehan, Julien O. Teitler, Kathryn M. Neckerman, Tanya K. Kaufman, Gina S. Lovasi, Michael D. M. Bader, and Andrew G. Rundle. Neighborhood physical disorder in New York City. *Journal of Maps*, 12(1):53–60, 1 2016.

[RBR+11] Andrew G Rundle, Michael D M Bader, Catherine A Richards, Kathryn M Neckerman, and Julien O Teitler. Using Google Street View to audit neighborhood environments. *American journal of preventive medicine*, 40(1):94–100, 1 2011.

[RDS+15] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 12 2015.

[RM07] Lior Rokach and Oded Maimon. *Data Mining with Decision Trees*, volume 69 of *Series in Machine Perception and Artificial Intelligence*. World Scientific, 2nd edition, 12 2007.

[ROB07] R. J. F. Rossetti, E. C. Oliveira, and A. L. C. Bazzan. Towards a specification of a framework for sustainable transportation analysis. In *13th Portuguese Conference on Artificial Intelligence, EPIA, Guimarães, Portugal*, pages 179–190. APPIA, 2007.

[RRC17] M. A. Ramalho, R. J. F. Rossetti, and N. Cacho. Towards an architecture for smart garbage collection in urban settings. In *2017 International Smart Cities Conference (ISC2)*, pages 1–6, Sept 2017.

[RSR15] F. Rebelo, C. Soares, and R. J. F. Rossetti. Twitterjam: Identification of mobility patterns in urban centers based on tweets. In *2015 IEEE First International Smart Cities Conference (ISC2)*, pages 1–6, Oct 2015.

[SR17] Robert J Sampson and Stephen W Raudenbush. Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods Systematic Social Observation of Public Spaces: A New Look at Disorder in Urban Neighborhoods 1. 2017.

[SS01] L. G. Shapiro and C. G. Stockman. *Computer Vision*, volume 2. Pearson, 2001.

[TLFT11] Albert Torrent, Xavier Llado, Jordi Freixenet, and Antonio Torralba. Simultaneous detection and segmentation for generic objects. In *2011 18th IEEE International Conference on Image Processing*, pages 653–656. IEEE, 9 2011.

[TM08] Tinne Tuytelaars and Krystian Mikolajczyk. Local Invariant Feature Detectors: A Survey. *Computer Graphics and Vision*, 3(3):177–280, 2008.

[TP12] My T. Thai and P. M. (Panos M.) Pardalos. *Handbook of optimization in complex networks communication and social networks*. Springer Science+Business Media, LLC, 2012.

[TR96] John N Tsitsiklis and Benjamin Van Roy. Feature-Based Methods for Large Scale Dynamic Programming. *Machine Learning*, 22:59–94, 1996.

[Tri17] Bambang Trisakti. Vegetation type classification and vegetation cover percentage estimation in urban green zone using pleiades imagery. *IOP Conference Series: Earth and Environmental Science*, 54(1):012003, 1 2017.

[Tsa12] Chih-Fong Tsai. Bag-of-Words Representation in Image Annotation: A Review. *ISRN Artificial Intelligence*, 2012:1–19, 11 2012.

[Uni07] United Nations. Indicators of Sustainable Development : Guidelines and Methodologies. *New York*, (October):1–90, 2007.

[USRS16] D. Ulloa, P. Saleiro, R. J. F. Rossetti, and E. R. Silva. Mining social media for open innovation in transportation systems. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 169–174, Nov 2016.

[VD96] J.G. Verly and R.L. Delanoy. Model-based automatic target recognition (ATR) system for forwardlooking groundbased and airborne imaging laser radars (LADAR). *Proceedings of the IEEE*, 84(2):126–163, 2 1996.

[VKP+14] R. Veloso, Z. Kokkinogenis, L. S. Passos, G. Oliveira, R. J. F. Rossetti, and J. Gabriel. A platform for the design, simulation and development of quadcopter multi-agent systems. In *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6, June 2014.

[VOP+14] R. Veloso, G. Oliveira, L. S. Passos, Z. Kokkinogenis, R. J. F. Rossetti, and J. Gabriel. A symbiotic simulation platform for agent-based quadcopters. In *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6, June 2014.

[WP16] Hayden Wimmer and Loreen Powell. A Comparison of Open Source Tools for Data Science. *Journal of Information Systems Applied Research*, 9(2), 10 2016.

[WRL94] Bernard Widrow, David E. Rumelhart, and Michael A. Lehr. Neural networks: applications in industry, business and science. *Communications of the ACM*, 37(3):93–105, 3 1994.

[WW12] Jianguo Wu and Tong Wu. Sustainability indicators and indices: an overview. In *Handbook of Sustainability Management*, chapter 4, pages 65–86. World Scientific, London, 3 2012.

[YSSW17] S S Yu, Z C Sun, L Sun, and M F Wu. Extraction and Analysis of Mega Cities? Impervious Surface on Pixel-based and Object-oriented Support Vector Machine Classification Technology: A case of Bombay. *IOP Conference Series: Earth and Environmental Science*, 57(1):012042, 2 2017.

[YWT17] Haitao Yuan, Shuai Wang, and Jizong Tan. Study on municipal road cracking and surface deformation based on image recognition. In *AIP Conference Proceedings*, volume 1839, page 020004. AIP Publishing LLC, 5 2017.

[ZF14] Matthew D. Zeiler and Rob Fergus. Visualizing and Understanding Convolutional Networks. pages 818–833. Springer, Cham, 2014.

[ZGFD17] Xinyi Zhou, Wei Gong, WenLong Fu, and Fengtong Du. Application of deep learning in object detection. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pages 631–634. IEEE, 5 2017.

[ZRC14] A. Zaiat, R. J. F. Rossetti, and R. J. S. Coelho. Towards an integrated multi-modal transportation dashboard. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 145–150, Oct 2014.

REFERENCES

Appendix A

Table A.1: Research Questions Queries

	Search Query
RQ1	(Urbanization OR Urban Indicators) AND (Computer Vision OR Image Recognition)
RQ1	(Urbanization OR Urban Indicators) AND (Computer Vision OR Image Recognition) AND Survey
RQ1.1	(Urban Indicators OR Urban Planning OR Urbanization OR Landscape Metrics) AND (Definition OR Classification)
RQ1.2	(Urbanization OR Urban Indicators OR Urban Planning) AND (Classification OR categorization OR Labelling OR Categorizing OR Taxonomy)
RQ1.3	(Urban Indicators OR Urbanization) AND (Issues OR Challenges) AND (Retrieval OR Extraction)
RQ1.3	(Urban Indicators OR Urban Planning OR Urbanization OR Landscape Metrics) AND (Issues OR Challenges)
RQ1.3	(Urbanization OR Urban Indicators OR Progress measurement OR Urban Planning) AND Visual AND (Retrieval OR Retrieve)
RQ1.4	(Urbanization OR Urban Indicators OR Progress measurement OR Urban Planning) AND Visual
RQ1.4	(Urbanization OR Urban Indicators OR Progress measurement OR Urban Planning) NOT visual
RQ2	(Urbanization OR Urban Indicators) AND (Google Street View OR GSV OR Image Recognition API)
RQ2	(Urbanism OR Urbanization OR Disorder)AND (Vision API OR Image Recognition API)
RQ2.2	"Feature Extraction" AND Computer Vision
RQ2.3	"Image Description" AND Computer Vision
RQ2.4	"Image Description" AND Algorithm* AND Survey
RQ2.5	(Application OR API OR Google Street View) AND "Feature Extraction"
RQ2.6	(Issue OR Problem OR challenge) AND (Outdoor OR Real-life) AND Computer Vision AND Survey
RQ2.6	(Issue OR Problem OR challenge) AND (Outdoor OR Real-life) AND "Feature Extraction" AND Survey
RQ3	"visual semantic" AND "Urban Indicators"